STRENGTHENING EVIDENCE-SUPPORT SYSTEMS IN A CHALLENGED WORLD

HOW TO STRENGTHEN EVIDENCE-SUPPORT SYSTEMS BY FACILITATING THE CONNECTION BETWEEN EVIDENCE PRODUCERS AND USERS AND OPTIMIZING THE ROLE OF LIVING EVIDENCE SYNTHESES

By CRISTIÁN MANSILLA, M. Eng. Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2023) Hamilton, Ontario (Health Research Methods, Evidence, and Impact)

TITLE: How to strengthen evidence-support systems by facilitating the connection between evidence producers and users and optimizing the role of living evidence syntheses? AUTHOR: Cristián Mansilla M. Eng. Sc. SUPERVISOR: Professor John N. Lavis NUMBER OF PAGES: xiii, 226

Lay Abstract

Multiple insights can be considered when making policy and practice decisions. Here, we consider how evidence is used to make decisions, which is a topic that has likely never had the visibility it has had over the last several years. This dissertation addresses two key issues: (1) matching decision makers' needs with the right forms of evidence; and (2) ensuring that up-to-date summaries of the evidence evolve as contexts, issues and evidence evolve.

By using a cross-sectional survey, an online Delphi, and a critical interpretive synthesis this thesis offers ways to align evidence demand with supply, and to understand the role of living evidence syntheses in decision-making.

Three main outputs are produced: 1) a demand-driven taxonomy of the types of research questions; 2) a list of study designs that best address each question; and 3) a framework to understand what constitutes a "living evidence synthesis" and how and when to update them.

Abstract

The COVID-19 pandemic stimulated new innovations in evidence support and exacerbated long-standing challenges confronting those providing evidence support. To build stronger and more sustainable evidence-support systems, two key issues are important: 1) matching a decision maker's need to the right combination of forms of evidence; and 2) ensuring updated summaries of the evidence are available when decision-makers need it. This dissertation aims to address both issues by (1) creating a taxonomy of demand-driven types of question for which research could provide insight; (2) building lists of study designs that optimally address each type of question; and (3) producing a theoretical framework to better understand what constitutes a living evidence synthesis, when and how to update them, and their role in the decision-making process.

The first study is a cross-sectional survey targeting units providing evidence support to decision makers to create a demand-driven taxonomy of types of question. The second study is an online Delphi process asking methodological experts to create a list of study designs to answer these questions. Finally, study 3 is a critical interpretive synthesis to create a theoretical framework to understand living evidence syntheses and their role in decision-making processes.

In chapter 2, 29 participants responded the cross-sectional survey, and a taxonomy of 40 demand-driven types of questions structured in the four main decision-making stages was created. In chapter 3, 29 methodological experts participated in the online Delphi process, and consensus was reached for 28 out of the 40 types of questions. Finally, in chapter 4, 152

publications were included, and six thematic categories were found to produce a conceptual framework.

Together, the first two studies provide a way to facilitate the alignment between evidence demand and supply, while the third study helps to clarify the role of living evidence syntheses in decision-making processes.

Acknowledgments

I started this journey on August 15th of 2019 by jumping on a plane from Santiago to Pearson Airport with my wife and my 1.5 years-old daughter. I never thought that I would finish this process 4 years later, with a 5-year-old and 2-months old daughters. This was an incredible journey with many people to acknowledge. Here are the most relevant ones with some explicit mentions in Spanish!

First and foremost, to my family. My life-partner Daniela, my oldest daughter Isabel, and my recently born Elena. Dani, gracias por tu incondicionalidad y la gran apuesta que hiciste al querer venirte conmigo a estas tierras, y compartir la vida y nuestra familia juntos en una latitud lejana. Nada de esto habría sido igual si no hubieses estado tú a nuestro lado. A la Isabel, parece increíble que un día de Septiembre de ese mismo 2019 te dejamos en el jardín de la Universidad con niños y profesoras que hablaban un idioma que nunca habías escuchado y ahora, con ya 5 años, cantas el himno de Canadá con orgullo, y disfrutas cada temporada del año con todas sus actividades. Gracias por apañar a todos y cada uno de los panoramas de tu papá y por seguir compartiendo toda tu alegría y espontaneidad. Y, por último, a la Elena, nuestra guagüita canadiense. Aunque aún tengas unos poquitos meses de vida, mantengo la imagen de tu incipiente sonrisa en cada momento del día. Gracias por acompañar a toda tu familia, por iluminar la vida de tus padres y la de tu hermana. Ustedes son parte integral de haber culminado este proceso.

To my supervisor John Lavis. Thank you for your generous and humble style of supervision. I still remember when we met in Santiago back in 2018, and I talked to you about my intentions to do a PhD, and you gave me a hundred names of potential supervisors that would match my interests, without ever convincing me that you could be a good choice. And, despite your vast experience, I always felt that my ideas were taken into consideration throughout all my work at the Forum and my dissertation. I hope our fruitful relationship could bring many other great outputs.

To my committee members, Arthur Sweetman and Gordon Guyatt. I remember John said to me once “You have two of the brightest minds of the University in your committee, you should record your committee meetings!”. He was absolutely right. Thank you for your critical insights and time devoted to every single part of the thesis.

The COVID-19 pandemic deserves a special mention. I have never thought that I would live such difficult times with no family support and in a different country. Un gran reconocimiento a la Dani con quien vivimos juntos esos tiempos difíciles, y a la Isabel por haber sorteado con gran valentía ese tiempo, teniendo tan solo 2 añitos.

To my extended family, my parents, and siblings. A pesar de que no estuvieron físicamente conmigo durante este proceso, fueron siempre refugio de energía y cariño, tanto para mí, como para mis ahora 2 hijas. A mis papás Manuel y Magdalena, mis hermanos Cata y Nico, y a mi familia política Gloria y Patricio, y al Pablo. Gracias por acompañarme a mí y a toda nuestra familia en este viaje.

To my cohort, Kerry, Jeonghwa, Tatiana and Puspita, for all our time together sharing PhD life, comps, and making the post-pandemic PhD experience something real.

To my friends (almost family) in Canada. En particular a los amigos chilenos de la Fonda y los latinos que ya casi son chilenos. Gracias por tantos asados, paseos y salidas juntos.

Finally, thank you for the financial support that I received to pursue this PhD. The Queen Elizabeth Strengthening Health & Social Systems Scholarship, and the ANID Becas Chile para Doctorado en el Extranjero support from the Government of Chile.

Thank you, everyone, for joining me on this journey!

Table of contents

List of figures and tables

Chapter 2. A taxonomy of demand-driven questions for use by evidence producers, intermediaries, and decision-makers: results from a cross-sectional survey

Chapter 3: Matching the right study design to decision-maker questions: results from a Delphi study.

Chapter 4: A living critical interpretive synthesis on the production and dissemination of living evidence syntheses for decision making.

List of abbreviations

CIS: Critical interpretive synthesis

Evidence-informed health policymaking (EIHPM)

KTP: Knowledge Translation Platform

KW: Kendall's W

LES: Living evidence synthesis

LMIC: Low-and-middle income country

RC: Rank correlation

WHO: World Health Organization

Declaration of academic achievement

This dissertation includes three chapters of original research (Chapters 2, 3, and 4), in addition to an introduction and conclusion (Chapters 1 and 5, respectively). I am the lead author of each of the three original research chapters. I conceptualized the overall dissertation including the focus and design of each original thesis chapter; designed data collection instruments and undertook data collection; conducted analysis; and prepared written manuscripts. My supervisor Dr. John Lavis and committee members Dr. Arthur Sweetman and Dr. Gordon Guyatt provided input into the study design, development of study materials, and analysis of data, and provided feedback on drafts. They were co-authors of all of my papers. Additionally, Chapter 4 involved a number of additional co-authors who helped me in the process of selecting and extracting data from the articles that were included in the critical interpretive syntheses. They are Qi Wang, Thomas Piggott, Kerry Waddell, and Peter Bragge.

Chapter 1: Introduction

1.1.Overview of the chapter

This chapter is structured in six main sections. First, it presents the general background on the topic that this thesis addresses by providing critical insights on evidence-support systems before, during and after the COVID-19 pandemic, and by identifying key areas where the field will need to evolve for future global crises. Secondly, the chapter introduces the general and specific objectives of the thesis and outlines how the methodology used addresses each of the objectives. Finally, this chapter foreshadows the anticipated contributions of this dissertation to the field by reflecting on the creation of a methodological tool, as well as a theoretical framework, and introducing the rest of the chapters.

1.2.Background

Almost 20 years ago, the World Health Organization (WHO) included in its 2004 World Report on Knowledge for Better Health a chapter dedicated to 'linking research to action' (World Health Organization, 2004), which was followed by a 2005 World Health Assembly resolution encouraging countries to "establish or strengthen mechanisms of knowledge transfer to support public health development, health-related policies and evidence-based health systems" (World Health Assembly, 2005). Later (and following a number of subsequent events, including the 2008 Bamako Call to Action on Research for Health (The Lancet, 2008), and a critical assessment on how much the World Health Organization was considering research evidence when making recommendations (A. D. Oxman et al., 2007)), these calls came to be seen as the

beginning of evidence-informed health policymaking (EIHPM) as a new field focused on supporting the use of scientific evidence in health policymaking processes.

EIHPM is an approach that aims to ensure that decisions in health policymaking are based on the best available evidence (A. Oxman et al., 2009). This approach promotes the use of research evidence as a key input in the policymaking process, which could contribute to making better decisions and better policies. In 2009, the SUPPORT Tools provided a comprehensive framework on how to better enhance EIHPM, including 18 publications that provide guidance on how to support evidence-informed policymaking, how to identify evidence needs in the policymaking process, how to find and assess different types of evidence to inform decisions, and how to navigate from research evidence to making decisions (Lavis, Oxman, et al., 2009). A key challenge for EIHPM was also very clear from the beginning, in terms of how can evidence ecosystems bridge the gap between the availability of scientific knowledge and its uses across different levels of the health system, including the policymaking process (Oliver et al., 2014; Partridge et al., 2020).

Later again, knowledge translation platforms (KTPs) emerged as a critical strategy to create and institutionalize country-level efforts to link research to action by facilitating partnerships between policymakers, researchers, civil society organizations and other stakeholders and thus promote the use of evidence in decision making processes (Lavis et al., 2006). The development and implementation of multiple KTPs all over the world (Adu et al., 2021; Berman et al., 2015; Kasonde & Campbell, 2012; Mansilla et al., 2017; Ongolo-Zogo et al., n.d.) – many of them created with the sponsorship of the WHO Evidence-informed Policymaking Network (EVIPNet)

initiative (Hamid et al., 2005; World Health Organization, 2016) – facilitated both domestic policy development and implementation through the use of the best available scientific evidence, and the connection of country-teams with regional and global bodies to facilitate drawing lessons across countries and regions.

Beyond KTPs, some domestic efforts to promote EIHPM across countries have been undertaken, but there is still a strong need for coordinated global support efforts. These initiatives have often been isolated and lacked coordination with other countries and regions, the consequence of which is that important lessons and promising practices are not being shared, and that many countries and governments are not benefiting from the lessons learned in other countries when encouraging EIHPM efforts. Such a consequence may be particularly harmful for low- and middle-income countries (LMICs), which often lack the resources and infrastructure to implement evidence-informed policies and programs.

There are a number of global coordination efforts to convene evidence producers (e.g., researchers) – such as Cochrane (Chalmers, 1993), Campbell (Petrosino et al., 2001) and the Guidelines International Network (GIN) (Ollenschlager, 2004) – but very little has been done in terms of how to connect the demand for evidence (i.e., coming from the multiple needs that decision makers commonly have) with the full array of types of evidence producers.

The COVID-19 pandemic created an unprecedented challenge for decision making processes, bringing to attention the importance of evidence-informed decision making, and catalyzing important global efforts to support it (Khalil et al., 2022; Vickery et al., 2022). During this global

crisis, decision makers needed urgent access to timely, reliable, and relevant scientific evidence to make multiple decisions that had significant consequences for population and global health.

The latest reports from the WHO show that more than 6 million people have died and more than 700 million have had a confirmed cases of COVID-19 around the world (World Health Organization, 2023). The COVID-19 pandemic has been the most important global health event of this century, having dramatic consequences to world population health, but also to the evidence system, which has seen an explosion of data and research produced.

The production of research in the last two decades had escalated to unprecedented levels, not only at the level of primary studies, but also with an increase production of evidence syntheses. Thus, and due to the massive production of research, the long-recommended strategy of using evidence syntheses instead of individual primary studies to inform decisions has become even more important, and more challenging, since for any given topic, a critical analysis of the best available evidence syntheses is needed.

This massive production of research was substantially accelerated during the COVID-19 pandemic (Kambhampati et al., 2020), creating important new challenges such as potential duplication and making it difficult to keep abreast of the most recent decision-relevant evidence (Pearson, 2021). The massive production of research during the COVID-19 pandemic has not only been seen among researchers producing primary studies, but also at the level of evidence syntheses. In November 2021 (only 20 months after it was declared a pandemic by the WHO), more than 8,000 evidence syntheses had been produced, with a disproportionate number

addressing clinical management issues, as opposed to other health system, economic and social issues related to the pandemic (unpublished work), and having significant areas of duplication (Beresford et al., 2022).

Moreover, the COVID-19 pandemic not only created a unique moment for EIHPM by challenging evidence demand (i.e., multiple, constantly evolving, and urgent decision-making needs) and evidence supply (i.e., high-volume evidence production), but it also brought an opportunity to truly think about decision-making in an intersectoral way. In fact, the global crisis produced due to the COVID-19 pandemic inevitably involved multiple intersectoral issues that could not be addressed by only one sector.

This dissertation will address two key issues that are critical to any evidence-support system and will reflect on them trying to address key considerations on how to further advance efforts to strengthen the global evidence architecture. First, how can we facilitate the matching between a decision maker's need and the right combination of forms of evidence to address this issue. Secondly, how to keep abreast with the constantly evolving nature of the evidence on a given topic, in a context that is moving fast and constantly changing.

Despite the massive production of evidence during the most acute phase of the COVID-19 pandemic, the topic coverage of the research produced was uneven (unpublished work) and the right mix of forms of evidence were not always used to address different decision makers' needs (Global Commission on Evidence to Address Societal Challenges, 2022). While policymakers often tend to rely on data analytics and evaluations to inform their decisions, there is a lack of a

unified evidence-support system that could address different needs with the right combination of the different forms of evidence.

In recent years, there has been a growing number of global calls to better coordinate and strengthen the global evidence infrastructure (Cochrane Convenes, 2022; Global Commission on Evidence to Address Societal Challenges, 2023; Kuchenmüller et al., 2022). These calls have raised the issue that there is a critical need to properly match and integrate the different forms of evidence to support the steps and varied needs throughout the decision-making process.

These calls are made in a context in which decision-making processes are understood at multiple levels and with a variety of decision makers. In this context, multiple approaches and strategies have promoted the use of evidence in decision-making processes (Graham et al., 2006; Ward et al., 2009; World Health Organization, 2022), highlighting the importance of this issue as a critical step in making better policies.

Here, we can refer to decision makers as government policymakers, organizational leaders, professionals, and citizens. Additionally, we understand an evidence-support system as structures where decisions are made by decision-makers, and they seek advice from people filling multiple roles. Here, we could have people that provide advice for a specific topic, or people providing advice from a more methodological perspective.

Despite the potential benefits of having a strong evidence-support system that could bring together different forms of evidence, efforts to establish such a system have been scattered. However, as the COVID-19 pandemic has brought significant attention to the importance of using evidence to make policy decisions, there is an invaluable opportunity to further strengthen evidence-support systems, and to do so in ways that bring together different forms of evidence, depending on the specific decision-making need at a given point in time. A strong evidence-support system addresses decision-makers' questions with the right mix of forms of evidence, which entails matching the right combination of form of evidence to the specific question that a decision maker might have, depending on the step in the decision-making process in which this question arises.

Another challenge that the COVID-19 pandemic brought was the change in the nature of evidence demand. While policymakers often had a number of different needs that required evidence support in a quick turnaround time (Ganann et al., 2010), the contexts in which policy decisions were taking place were shifting significantly faster than in normal times. This meant that decision-makers needed rapid advice on multiple questions to address complex decisions in rapidly changing contexts (Vickery et al., 2022).

Here, multiple innovations were developed or strengthened during the COVID-19 pandemic to address this issue. For example, rapid reviews that already provided support in a shorter turnaround period of time, compared to systematic reviews (Tsertsvadze et al., 2015), were further accelerated to provide even faster support to address the more urgent needs of decision-makers (Rehfuess et al., 2022; Tricco et al., 2022).

Another critical development in this area was the further development of the concept of living evidence syntheses (LESs), which provide continuously updated summaries of an evolving evidence base. Although LESs were first described before the COVID-19 pandemic (Elliott et al., 2014, 2017), they were produced on a larger scale during the COVID-19 pandemic. This can be explained in response to the evolving nature of evidence (and the need of keeping abreast of the most up-to-date evidence in a given topic) as well as the evolving contexts and issues for which the evidence was being summarized during these troubled times.

LESs emerged as a key strategy to address the challenges posed by the COVID-19 pandemic, and they are now considered a key cornerstone of evidence infrastructure (Akl et al., 2020). An investment in an evolving suite of LESs in key areas where decision makers have the most pressing evidence needs could produce a stronger evidence-support system, particularly during global crises, where the demand for and supply of evidence evolve at a faster speed.

1.3.Aims and objectives.

This dissertation aims to provide a contribution to domestic (e.g., national, or sub-national) evidence-support systems and the global evidence architecture by creating two frameworks to facilitate question formulation and the selection of study designs, as well as the role of LESs for decision making. Specifically, this thesis aims to:

1. create a taxonomy of demand-driven types of question that act as a mutually exclusive and collectively exhaustive list of questions for which research could provide insight.

2. build a list of study designs that optimally address each one of the types of question that are part of this taxonomy; and
3. produce a theoretical framework to clarify what LESs are, when and how to update them, and their role in the decision-making process.

1.4.Design and methodology of the thesis

This thesis includes three studies that address the three objectives outlined above. Each study uses different methods to achieve its objectives.

The first study is a cross-sectional survey that was conducted to collect a broad range of decision-makers' questions by asking different units their evidence-informed answers. This global survey targeted units that provide some type of evidence support (e.g., evidence syntheses) to decision makers. Using this pool of questions, an iterative analysis was conducted to create a demand-driven taxonomy of types of questions, thereby creating a mutually exclusive and collectively exhaustive list of questions, and one that is organized in the four main decision-making stages (i.e., clarifying a societal problem, its causes and potential impacts; finding and selecting options to address a problem; implementing or scaling-up an option; monitoring implementation and evaluating impacts).

The second study builds on the first and used an online Delphi process to create a list of study designs that are most suitable to answer each type of question that was included in the taxonomy created in study 1. Here, a group of experts in methodology were invited to participate in a two-round Delphi in which they were asked to rank a list of study designs in terms of their potential

and suitability to answer each type of question. The level of consensus achieved in each round was assessed and a list of study designs per type of question was presented as a key output of this study.

Finally, the third study creates a theoretical framework to understand what are LESs and when should we produce them. This study used a critical interpretive synthesis (CIS) method, which is a type of evidence synthesis that aims to produce a conceptual framework from the critical interpretation of themes emerging from the literature on a given topic.

1.5.Anticipated contributions of the thesis

This thesis contributes to address two key areas in the future of evidence-support systems, namely, how to better connect the different forms of evidence with the specific needs that decision makers might have, and to better understand the role that LESs could have in decision making, especially for rapidly evolving contexts, such as global crises.

The first two studies will create a novel tool that will facilitate a better connection of decision makers' needs with different study designs (and, hence, to the different forms of evidence) that might provide a more suitable answer to a specific need. On the one hand, this tool will facilitate the work of evidence intermediaries to help them to better connect evidence demand (i.e., what decision makers ask) with evidence supply (i.e., what researchers and evidence producers create). On the other hand, this tool will facilitate the work of decision makers by helping them ask more specific questions, thus using the taxonomy to understand the types of questions from which they might expect to receive insights from evidence. Finally, this tool can also help

evidence producers and researchers to better choose the study designs to answer specific needs depending on the context.

The third study will provide important contributions by creating a framework that will facilitate the understanding of the production and use of LESs for decision making during both future global crises and 'regular' times. This conceptual framework will build on the existing literature about LESs that mainly sets the scene on what LESs are, and will provide a more in-depth definition of LESs, as well as critical views on when they can be produced (or updated, or discontinued), and what methods we have to make the findings of LESs available.

1.6.Structure of the thesis

Besides this chapter, this thesis is structured in four main chapters. Chapter 2 presents the result of study 1 which, as described above, used a global survey to create a taxonomy of mutually exclusive and collectively exhaustive list of types of questions. Chapter 3 builds on chapter 2 and creates a list of study designs that are preferable to address each one of the types of question that were identified in the previous study. Chapter 4 describes the results of study 3, which used a CIS to produce a conceptual framework that explains what LESs are, how they are conducted, when they need to be produced or updated, and ways to disseminate their findings. Finally, Chapter 5 is the conclusion, describing how this thesis contributes to advance knowledge in this area, as well as key implications for policy and practice and for future research.

1.7.References

Adu, J., Gyamfi, S., & Martin-Yeboah, E. (2021). Knowledge translation platforms to support African evidence-informed policies: Challenges and progress. Evidence & Policy: A Journal of Research, Debate and Practice. https://doi.org/10.1332/174426421X16123456824061

Akl, E. A., Haddaway, N. R., Rada, G., & Lotfi, T. (2020). Future of Evidence Ecosystem Series: Evidence synthesis 2.0: When systematic, scoping, rapid, living, and overviews of reviews come together. Journal of Clinical Epidemiology, 123(PG-162-165), 162–165. https://dx.doi.org/10.1016/j.jclinepi.2020.01.025

Beresford, L., Walker, R., & Stewart, L. (2022). Extent and nature of duplication in PROSPERO using COVID-19-related registrations: A retrospective investigation and survey. BMJ Open, 12(12), e061862. https://doi.org/10.1136/bmjopen-2022-061862

Berman, J., Mitambo, C., Matanje-Mwagomba, B., Khan, S., Kachimanga, C., Wroe, E., Mwape, L., van Oosterhout, J. J., Chindebvu, G., van Schoor, V., Ritchie, L. M. P., Panisset, U., & Kathyola, D. (2015). Building a knowledge translation platform in Malawi to support evidence-informed health policy. HEALTH RESEARCH POLICY AND SYSTEMS, 13.

Chalmers, I. (1993). The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care. Annals of the New York Academy of Sciences, 703(1 Doing More Go), 156–165. https://doi.org/10.1111/j.1749-6632.1993.tb26345.x

Cochrane Convenes. (2022). Preparing for and responding to global health emergencies: Learnings from the COVID-19 evidence response and recommendations for the future. February 2022. https://doi.org/10.6084/m9.figshare.19115849.v2

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., … Pearson, L. (2017). Living systematic review: 1. Introduction—The why, what, when, and how. Journal of Clinical Epidemiology, 91, 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., & Gruen, R. L. (2014). Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. PLoS Medicine, 11(2 PG-e1001603), e1001603. https://dx.doi.org/10.1371/journal.pmed.1001603

Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: Methods and implications of rapid reviews. Implementation Science, 5(1), 56. https://doi.org/10.1186/1748-5908-5-56

Global Commission on Evidence to Address Societal Challenges. (2022). The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers. https://www.mcmasterforum.org/docs/default-source/evidence-commission/evidence-commission-report.pdf?Status=Master&sfvrsn=2fb92517_5/Evidence-Commission-report

Global Commission on Evidence to Address Societal Challenges. (2023). Evidence Commission update 2023: Strengthening domestic evidence-support systems, enhancing the global evidence architecture, and putting evidence at the centre of everyday life. McMaster Health Forum. https://www.mcmasterforum.org/docs/default-source/evidence-commission/update-2023.pdf?sfvrsn=e81cbf_8

Hamid, M., Bustamante-Manaog, T., Dung, T. V., Akkhavong, K., Fu, H., Ma, Y., Zhong, X., Salmela, R., Panisset, U., & Pang, T. (2005). EVIPNet: Translating the spirit of Mexico. The Lancet, 366(9499), 1758–1760. https://doi.org/10.1016/S0140-6736(05)67709-4

Kambhampati, S. B. S., Vaishya, R., & Vaish, A. (2020). Unprecedented surge in publications related to COVID-19 in the first three months of pandemic: A bibliometric analytic report. Journal of Clinical Orthopaedics and Trauma, 11, S304–S306. https://doi.org/10.1016/j.jcot.2020.04.030

Kasonde, J. M., & Campbell, S. (2012). Creating a Knowledge Translation Platform: Nine lessons from the Zambia Forum for Health Research. Health Research Policy and Systems, 10(1), 31. https://doi.org/10.1186/1478-4505-10-31

Khalil, H., Tamara, L., Rada, G., & Akl, E. A. (2022). Challenges of evidence synthesis during the 2020 COVID pandemic: A scoping review. Journal of Clinical Epidemiology, 142, 10–18. https://doi.org/10.1016/J.JCLINEPI.2021.10.017

Kuchenmüller, T., Lavis, J., Kheirandish, M., Reveiz, L., Reinap, M., Okeibunor, J., Siswanto, S., Rashidian, A., Sieber, S., Moat, K., Mansilla, C., El-Jardali, F., Helble, M., Reeder, J., Chapman, E., Barreto, J. O. M., Mandil, A., & Swaminathan, S. (2022). Time for a new global roadmap for supporting evidence into action. PLOS Global Public Health, 2(6), e0000677. https://doi.org/10.1371/journal.pgph.0000677

Lavis, J. N., Lomas, J., Hamid, M., & Sewankambo, N. K. (2006). Assessing country-level efforts to link research to action. Bulletin of the World Health Organization, 84(8), 620–628.

Lavis, J. N., Oxman, A. D., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP). Health Research Policy and Systems, 7(S1), I1, 1478-4505-7-S1-I1. https://doi.org/10.1186/1478-4505-7-S1-I1

Mansilla, C., Herrera, C. A., Basagoitia, A., & Pantoja, T. (2017). The Evidence-Informed Policy Network (EVIPNet): Lessons learnt from 1 year of coordinated efforts in evidence informed policymaking in Chile. Pan American Journal of Public Health, 41(e36).

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. BMC Health Services Research, 14(1), 2. https://doi.org/10.1186/1472-6963-14-2

Ollenschlager, G. (2004). Improving the quality of health care: Using international collaboration to inform guideline programmes by founding the Guidelines International Network (G-I-N). Quality and Safety in Health Care, 13(6), 455–460. https://doi.org/10.1136/qshc.2003.009761

Ongolo-Zogo, P., Lavis, J. N., Tomson, G., & Sewankambo, N. K. (n.d.). Assessing the influence of knowledge translation platforms on health system policy processes to achieve the health millennium development goals in Cameroon and Uganda: A comparative case study. https://doi.org/10.1093/heapol/czx194

Oxman, A. D., Lavis, J. N., & Fretheim, A. (2007). Use of evidence in WHO recommendations. The Lancet, 369(9576), 1883–1889. https://doi.org/10.1016/S0140-6736(07)60675-8

Oxman, A., Lavis, J., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 1: What is evidence-informed policymaking? Health Research Policy and Systems, 7(Suppl 1), S1. https://doi.org/10.1186/1478-4505-7-S1-S1

Partridge, A. C. R., Mansilla, C., Randhawa, H., Lavis, J. N., El-Jardali, F., & Sewankambo, N. K. (2020). Lessons learned from descriptions and evaluations of knowledge translation platforms supporting evidence-informed policy-making in low- and middle-income countries: A systematic review. Health Research Policy and Systems, 18(1), 127. https://doi.org/10.1186/s12961-020-00626-5

Pearson, H. (2021). How COVID broke the evidence pipeline. Nature, 593(7858), 182–185. https://doi.org/10.1038/d41586-021-01246-x

Petrosino, A., Boruch, R. F., Soydan, H., Duggan, L., & Sanchez-Meca, J. (2001). Meeting the Challenges of Evidence-Based Policy: The Campbell Collaboration. The ANNALS of the American Academy of Political and Social Science, 578(1), 14–34. https://doi.org/10.1177/000271620157800102

Rehfuess, E. A., Burns, J. B., Pfadenhauer, L. M., Krishnaratne, S., Littlecott, H., Meerpohl, J. J., & Movsisyan, A. (2022). Lessons learnt: Undertaking rapid reviews on public health and social measures during a global pandemic. Research Synthesis Methods, 13(5), 558–572. https://doi.org/10.1002/jrsm.1580

The Lancet. (2008). The Bamako call to action: Research for health. The Lancet, 372(9653), 1855. https://doi.org/10.1016/S0140-6736(08)61789-4

Tricco, A. C., Straus, S. E., Ghaffar, A., & Langlois, E. V. (2022). Rapid reviews for health policy and systems decision-making: More important than ever before. Systematic Reviews, 11(1), 153. https://doi.org/10.1186/s13643-022-01887-7

Tsertsvadze, A., Chen, Y.-F., Moher, D., Sutcliffe, P., & McCarthy, N. (2015). How to conduct systematic reviews more expeditiously? Systematic Reviews, 4(1), 160. https://doi.org/10.1186/s13643-015-0147-7

Vickery, J., Atkinson, P., Lin, L., Rubin, O., Upshur, R., Yeoh, E.-K., Boyer, C., & Errett, N. A. (2022). Challenges to evidence-informed decision-making in the context of pandemics: Qualitative study of COVID-19 policy advisor perspectives. BMJ Global Health, 7(4), e008268. https://doi.org/10.1136/BMJGH-2021-008268

World Health Assembly. (2005). Resolution on health research 2005. https://apps.who.int/gb/ebwha/pdf_files/WHA58-REC1/english/A58_2005_REC1-en.pdf

World Health Organization. (2004). The world health report. 2004: Changing history. https://apps.who.int/iris/bitstream/handle/10665/42891/924156265X.pdf

World Health Organization. (2016). EVIPNet in action: 10 years, 10 stories. http://www.who.int/evidence/resources/publication/en/

World Health Organization. (2023). WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int

Chapter 2: A taxonomy of demand-driven questions for use by evidence producers, intermediaries, and decision-makers: results from a cross-sectional survey

2.1. Preface

There is a globally recognized need to formalize and strengthen evidence-support systems, which can contribute to inform decision-making processes. To do so, it is critical that evidence producers and decision-makers interact, so evidence needs can be easily formulated as questions that can be answered by evidence producers. This chapter presents a taxonomy of demand-driven questions, which was built by categorizing hundreds of demand-driven questions collected through a global cross-sectional survey of units providing some type of evidence support.

Dr John Lavis and I were responsible for conceiving the idea, and I designed the protocol, collected, and analyzed data, and wrote the manuscript. Dr. Arthur Sweetman and Dr. Gordon Guyatt provided critical insights on the study design, and particularly data analysis by ensuring to have a robust output. All co-authors provided feedback on the draft manuscript and the paper was submitted to Implementation Science on 31 July 2023.

A taxonomy of demand-driven questions for use by evidence producers, intermediaries, and decision-makers: results from a cross-sectional survey

Cristián Mansilla*[1,2], camansil@gmail.com

Arthur Sweetmann[3], arthur.sweetman@mcmaster.ca

Gordon Guyatt [4] guyatt@mcmaster.ca

John N Lavis[1,4] lavisj@mcmaster.ca

[1] McMaster Health Forum, McMaster University, 1280 Main St W MML-417, Hamilton, Ontario, Canada, L8S 4L6

[2] Health Policy PhD program, McMaster University, 1280 Main St W, Hamilton, Ontario, Canada, L8S 4L6

[3] Faculty of Social Sciences, McMaster University, 1280 Main St W Kenneth Taylor Hall Rm. 129, Hamilton, Ontario, Canada, L8S 4M4

[4] Department of Health Research Methods Evidence and Impact, McMaster University, 1280 Main St W 2C Area, Hamilton, Ontario, Canada, L8S 4K1

Word count: 2,872

2.2. Abstract

Background

Globally, a growing number of calls to formalize and strengthen evidence-support systems have been released, all of which emphasize the importance of evidence-informed decision making. To achieve this, it is critical that evidence producers and decision-makers interact, and that decision-makers' evidence needs can be efficiently translated into questions to which evidence producers can respond.

This paper aims to create a taxonomy of demand-driven questions for use by evidence producers, intermediaries (i.e., people working in between researchers and decision-makers) and decision-makers.

Methods

We conducted a global cross-sectional survey of units providing some type of evidence support at the explicit request of decision-makers. Unit representatives were invited to answer an online questionnaire where they were asked to provide a list of the questions that they have addressed through their evidence-support mechanism. Descriptive analyses were used to analyze the survey responses, while the questions collected from each unit were iteratively analyzed to create a mutually exclusive and collectively exhaustive list of types of questions that can be answered with some form of evidence.

Results

Twenty-nine individuals completed the questionnaire, and more than 250 submitted questions were analyzed to create a taxonomy of 40 different types of demand-driven types of questions. These 40 questions were organized by the goal to be achieved, and the goals were grouped in the four decision-making stages ((i) clarifying a societal problem, its causes, and potential impacts; (ii) finding and selecting options to address a problem; (iii) implementing or scaling-up an option; and (iv) monitoring implementation and evaluating impacts).

Conclusion

The mutually exclusive and collectively exhaustive list of demand-driven questions will help decision-makers (to ask and prioritize questions that evidence producers and intermediaries can answer), evidence producers (to organize and present their work), and evidence-intermediaries (to connect evidence needs with evidence supply).

KEYWORDS

Evidence-informed decision making; Evidence-informed policy; Evidence-based practice; Decision making; Methodology; Research questions.

2.3. Contributions to the literature

- Decision-makers' needs can often be translated into multiple questions that can be answered with research evidence.
- We present a taxonomy of demand-driven questions, which is a mutually exclusive and collectively exhaustive list of the types of questions where evidence could provide decision-relevant insights, organized by stage in the decision-making process.
- We collected decision-makers' questions addressed through some type of evidence-support mechanism in a global cross-sectional survey of evidence-support units.
- This taxonomy will help the evidence-support system by better connecting evidence demand and supply.

2.4. Background

Evidence has become a crucial component of decision-making processes and, by supporting decision-makers to address a broad variety of issues, from identifying problems to analyzing potential solutions and evaluating the implementation of actions, it can play a significant role in several stages of the policy cycle (Fretheim et al., 2009; Lavis, Wilson, Oxman, Grimshaw, et al., 2009; Lavis, Wilson, Oxman, Lewin, et al., 2009).

In recent years, there has been a growing number of calls to coordinate and strengthen the global evidence architecture (Cochrane Convenes, 2022; Global Commission on Evidence to Address Societal Challenges, 2022b; Kuchenmüller et al., 2022b). These calls stem from the recognition that evidence-informed decision making is essential for implementing better programs and policies, and that high-quality evidence is necessary for decision-making.

These calls have also stressed that there is a critical need to match and integrate the different forms of evidence to support the steps and varied needs in the decision-making process, and to further strengthen global evidence architecture. In this paper, we adopt the broad definition of evidence used by the Global Commission on Evidence to Address Societal Challenges (Global Commission on Evidence to Address Societal Challenges, 2022b), which includes all forms of decision-relevant evidence (data analytics, modelling, evaluation, qualitative insights, behavioral/implementation research, evidence syntheses, guidelines, and technology assessments).

Despite these global calls and the momentum created by the COVID-19 pandemic, there remains a continuing risk of mismatch between decision-makers' needs and the evidence that is made available to support decision-makers (Global Commission on Evidence to Address Societal Challenges, 2023). There are several factors that can help to explain why decision-makers' needs are not always fully addressed by research evidence (Brownson et al., 2006). One factor is that decision-makers have multiple evidence needs that are sometimes poorly served by the traditional ways that questions are organized by researchers (e.g., PICO (population, intervention, comparison, outcome) (Santos et al., 2007), SPIDER (sample, phenomenon of interest, design, evaluation, research type) (Cooke et al., 2012), and PEO (population, exposure, outcome)).

It is critically important that decision-makers understand what types of question that evidence might usefully address to, and that evidence producers and intermediaries (i.e., people working in between researchers and decision-makers) understand how to translate decision-makers' needs into the types of evidence that can be used to address these needs (Innvær et al., 2002). Such understanding can help to build trust, promote more and better interactions, and increase the usefulness and use of existing evidence.

This paper aims to create a taxonomy of questions that evidence can help to answer. Specifically, it aims to:

1. create a list of types of questions that decision-makers around the world have commonly asked of those they turn to for decision-relevant evidence; and

2. create a mutually exclusive and collectively exhaustive list of such questions.

2.5.Methods

This study is a cross-sectional survey of evidence-support units providing evidence support to decision-makers. These units provide evidence-related advice to decision makers on a timely and regular manner. The study aims to collect different types of questions that decision-makers regularly ask, to identify the wide range of questions where evidence could provide decision-relevant insights, and to develop a mutually exclusive and collectively exhaustive taxonomy of types of questions.

Participants

Between March and May 2022, representatives of evidence-support units were invited to answer a questionnaire, which was administered online via a link provided by email to each participant. To be eligible, units needed to:

- answer questions in response to a request coming from decision-makers, including (but not necessarily limited to) government policymakers (i.e., units addressing real-life evidence needs from decision-makers);
- address issues that are not exclusively in the clinical domain (for health-focused units); and
- have produced at least five evidence-informed answers in the last five years (i.e., the unit is or has recently been active).

Representatives of existing evidence-informed policymaking networks, the most recent of these being EVIPNet, were identified and contacted to verify if they were eligible to participate. Participants were also be asked if they were aware of other potentially eligible units.

Data collection

The online questionnaire requested the various types of questions that decision-makers regularly ask the unit and, when possible, for a more complete list of the questions they had previously addressed, a URL link to their products. The questionnaire also collected basic information regarding the scope of the work that each unit performs in supporting decision-making processes.

The questionnaire was originally written in English, but participants were also allowed to answer in French or Spanish if they felt more comfortable answering in those languages. The questionnaire is available in Additional File 1.

Data analysis

The data collected in the survey were summarised using descriptive analyses and reported with absolute numbers and frequencies. For each participant, the 10 most recent questions that each unit reported to have answered were collected.

Later, these questions were categorized in an iterative process to create a mutually exclusive and collectively exhaustive list. If necessary, compound questions answered by these units were split

into multiple fundamental questions, and questions were excluded if: (1) they were questions into which evidence cannot provide decision-relevant insights; (2) they were aiming to collect information about what other recommendations have said (e.g., what do scientific societies recommend about a given health condition?); (3) they were explicitly described as having not been asked by a decision-maker; and (4) they were addressed by building on other frameworks (e.g., agenda setting) that do not involve foreground evidence.

The initial draft taxonomy that was created from the responses was complemented by using existing frameworks included in the Evidence Commission report (Global Commission on Evidence to Address Societal Challenges, 2022), the GRADE Evidence to Decision (EtD) framework (Alonso-Coello et al., 2016), and the Consolidated Framework for Implementation Research (CFIR) (Damschroder et al., 2022). Finally, taking advantage of national, regional, and global meetings, a number of people were engaged in deliberations about how to improve the clarity and comprehensiveness of the taxonomy.

2.6.Results

Twenty-seven units were initially identified as potentially eligible, and seven additional units were suggested by participants. Two participants either declined or were found to be ineligible to participate, leaving 32 final potential participants. Twenty-nine answers were received (response rate 90.6%), but only 20 provided a list of questions that could be extracted. In total, 1076 questions were provided. By sampling the 10 most recent questions that were addressed by participants, we analyzed a total of 237 different questions.

Table 1 provides details about survey participants. The majority of the units surveyed were based in a university, national ministry, or non-governmental organization. While they accept requests from many types of actors, including government policymakers, managers, and program implementers, they most commonly answer requests coming from mid-level policymakers and program implementers. Finally, they serve different domains within the health sector, namely clinical management, public health decisions, health-system (not including technology assessment) decisions and technology assessments.

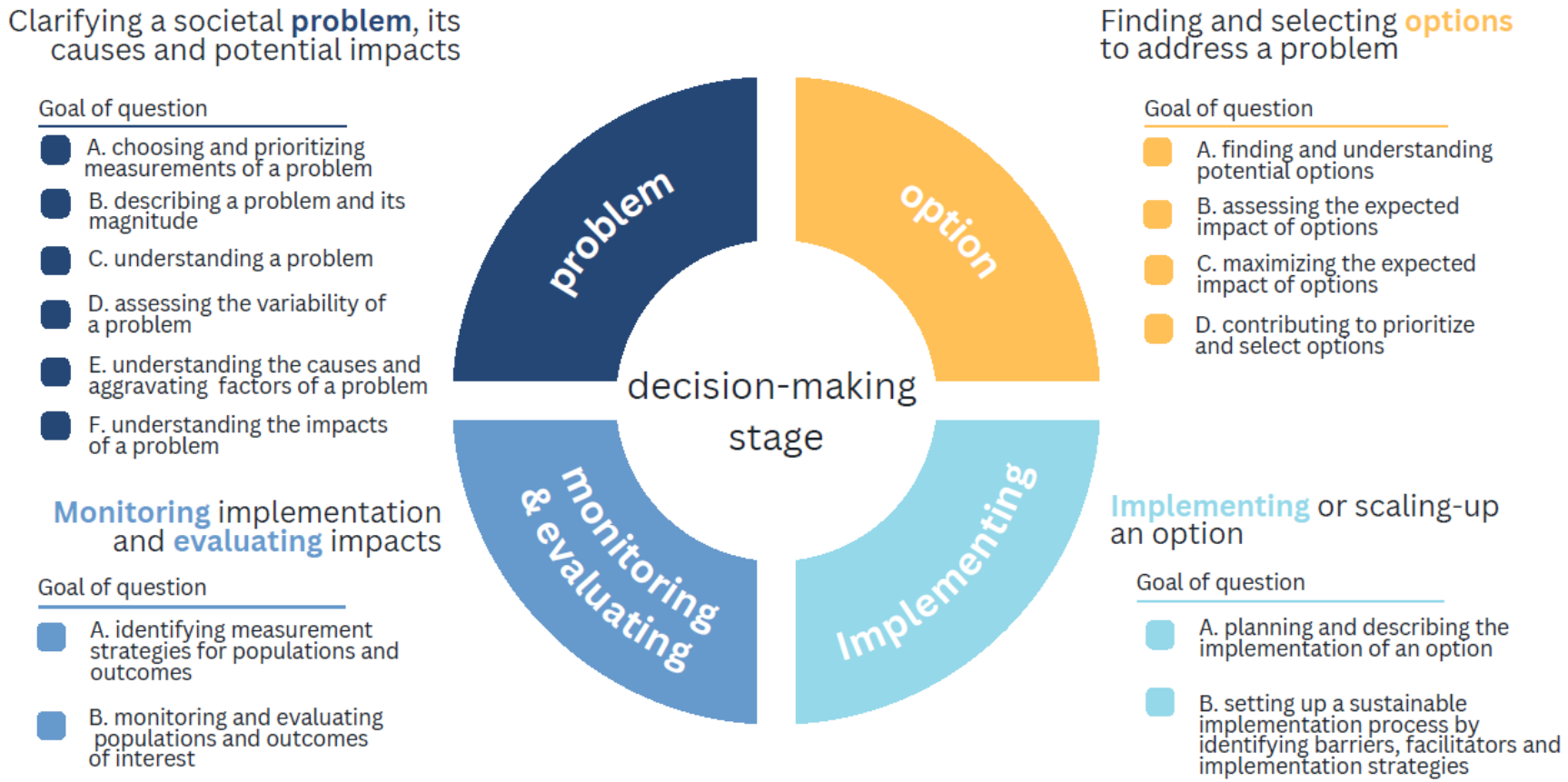
Table 2.1. Characteristics of units participating on the study

	N	%
Units currently active in answering decision-making needs		
Type of institution		
University	9	36%
National ministry	6	24%
Non-governmental organization	5	20%
Government agency	3	12%
Sub-national ministry	1	4%
Other	1	4%
Actors that are eligible to request evidence support		
Mid-level policymakers (e.g., head of units, departments)	24	96%
High-level policymakers (e.g., ministers, vice-ministers)	22	88%
Staff in charge of program implementation	21	84%
Managers in government agencies	18	72%
People working in NGOs	15	60%
People that are part of universities	12	48%
Other	4	16%
Actors that commonly request evidence support		
Mid-level policymakers (e.g., head of units, departments)	22	88%
Staff in charge of program implementation	16	64%
People working in NGOs	10	40%
High-level policymakers (e.g., ministers, vice-ministers)	9	36%
Managers in government agencies	9	36%
People that are part of universities	7	28%
Other	7	28%

Area and sector of work		
Public health	22	88%
Health systems (not including technology assessments)	21	84%
Clinical practice	12	48%
Health technology assessments	9	36%
Other	1	4%

Figure 1 shows the goals of each decision-making stage. In total, 41 different types of questions were identified and characterized as part of this taxonomy. To facilitate the understanding of the taxonomy, Tables 2 to 5 describe the types of questions included in each goal. A lay formulation of each goal is also provided in every table, and below. In each decision-making stage, to identify some concepts that are commonly used in certain disciplines to name specific types of questions, notes provide explanations of technical discipline-specific language. Additional File 2 presents a more detailed description of each type of question.

Figure 2.1. Taxonomy of demand-driven types of questions structured by decision-making stage.



Stage 1. Clarifying a societal problem, its causes, and potential impacts

This stage aims to clarify a problem, identify potential causes, and outline potential impacts or spillover effects that this problem might create. It is organized into six different goals that may need to be achieved (A to F). In total, this stage includes 15 different types of questions that may need to be answered (Table 2).

Table 2.2. Goals and types of questions for decision-making stage 1: clarifying a societal problem, its causes, and potential impacts.

Goals	Types of question
A. Choosing and prioritizing measurements of a problem. Lay language: how can a problem be measured?	1. Identifying measurements to characterize a problem
	2. Understanding individuals' values regarding outcomes
	3. Prioritizing measurements to characterize a problem
B. Describing a problem and its magnitude. Lay language: what's the problem and how big it is?	1. Describing a problem in a point in time
	2. Clarifying and characterizing populations affected by a problem
C. Understanding a problem. Lay language: how and why is a problem?	1. Finding conceptual approaches to understand a problem
	2. Understanding stakeholders' perceptions of a problem
	3. Understanding the role of context in a problem
D. Assessing the variability of a problem. Lay language: how the problem varies over time, across populations and in relation to other problems?	1. Assessing variability over time
	2. Assessing variability across populations and contexts
	3. Assessing the importance of a problem relative to other problems
E. Understanding the causes and aggravating factors of a problem. Lay language: what is causing or making the problem worse?	1. Identifying causes and/or aggravating factors of a problem
	2. Understanding the relative importance of causes and/or aggravating factors across populations and contexts

F. Understanding the impacts of a problem. Lay language: what impacts is the problem creating?	1. Identifying impacts/spillover effects of a problem
	2. Prioritizing the most important impacts/spillover effects of a problem

Technical language notes: In epidemiological research, describing a problem through frequencies is often called **prevalence** (e.g., number of people living with a given health condition) or **incidence** (e.g., number of people diagnosed with a given health condition during a certain time). In clinical research, the most common signs and symptoms of a given condition are often called the **clinical presentation** of a disease. In epidemiological research, causes can also be referred to as **risk or protective factors** that individuals can experience when they are **exposed** to a certain cause. In clinical research, the factors that could explain better or worse clinical outcomes on a given health condition are commonly called **prognostic factors**, while the potential causes of a health condition are called the **etiology of the disease**. In public health research, some potential factors that could explain different health outcomes are called **(social) determinants of health**. In some social sciences field, they could also be understood as **explanatory factors**, to understand what social factors would cause a given social behaviour. In economics, the unintended impacts of a given action are called externalities (e.g., passive smoking).

Although 'problems' create a decision-making scenario that frames an issue in a negative way, an issue can also be framed in a positive way as objectives (or once a problem has been identified, it can also be framed positively as objective). Then, the goals included in this section can also be framed in a positive or more neutral way by replacing problems by objectives, such as: A. Choosing and prioritizing measurements to determine whether an objective has been reached; B. Describing an objective and its implications; C. Understanding an objective; D. Assessing variability of an objective and its implications; E. understanding the preliminary steps and critical opportunities to reach out an objective; and F. Understanding the impacts of achieving an objective. We will continue by describing this stage as a 'problem' assuming that, as mentioned here, the question can be easily formulated using neutral or positive rhetoric.

Problems may be issues that are the in the present or the past, but they can also be issues that are not necessarily a problem now, but that could eventually become one (future problems, including

existential risk). These future problems were not created as specific types of questions, acknowledging that the same types of questions that are included in this stage can be equally formulated for future problems.

Problems can also arise from issues created in other decision-making stages (e.g., no feasible option is available, an implementation strategy does not address a barrier, or the option has not had the impact that it should have had, or its impact failed to be sustained). In these cases, users of this taxonomy might consider the issue as a new problem and identify a question that could match this issue in this decision-making stage.

Questions related to people's values and experiences (e.g., values regarding outcomes, understanding people's perceptions, etc.) might also vary according to some social characteristics, such as socioeconomic status, ethnicity, etc., and these issues are somehow included in these types of questions.

Stage 2. Finding and selecting options to address a problem.

This stage aims to find and select options that could address (or help to reduce) the impact of a problem. It is structured as four distinct goals that may need to be achieved (A to D). In total, this stage includes 13 different types of questions that may need to be answered (Table 3).

Table 2.3. Goals and types of questions for decision-making stage 2: finding and selecting options to address a problem.

Goals	Types of question
A. Finding and understanding potential options. Lay language: what are the potential solutions?	1. Scoping a list of potential options
	2. Understanding the way potential options and their components work
B. Assessing the expected impact of options. Lay language: is it feasible (can it work), does it work, is it convenient, and is it equitable and acceptable?	1. Assessing the feasibility of an option
	2. Assessing the benefits and early and frequently occurring harms of an option
	3. Identifying late-occurring harms of an option
	4. Assessing the acceptability of an option
	5. Assessing the costs and resource use of an option
	6. Assessing the efficiency in the use of resources
	7. Identifying equity, ethical, social, and human rights impact of an option
C. Maximizing the expected impact of options. Lay language: how can we ensure success with these solutions?	1. Adjusting options and enabling factors to maximize impact
	2. Finding population groups and contexts to focusing options
D. Contributing to prioritize and select options.	1. Creating packages of options
	2. Creating a ranking of options

Lay language: how to prioritize or combine solutions?	

Similar to problems, options can be present or past interventions, or they can also be interventions that are not available right now but could become an option in the future. Specific questions for these types of questions were not created, acknowledging that the same types of questions that are included in this stage can be formulated for present for future options.

The types of question included here are in the context of options not yet implemented and it is their possible impact that is assessed. The actual impact of the implementation of an option in decision-making will be addressed in stage 4 (Monitoring implementation and evaluating impacts).

Identifying the equity, ethical and human rights implications of an option could be understood as whether the impact of the option had different implications depending on specific population characteristics (e.g., socioeconomic status, ethnicity, etc.).

Stage 3. Implementing or scaling-up an option

This stage aims to address issues related to the implementation of a given option. It is structured around two different goals that may need to be achieved (A and B). In total, this stage includes 6 different types of questions that may need to be answered (Table 4).

Table 2.4. Goals and types of questions for decision-making stage 3: implementing or scaling-up an option.

Goals	Type of questions
A. Planning and describing the implementation of an option. Lay language: can it be done and what needs to happen to implement?	1. Identifying who has to do what to implement an option
	2. Identifying the context in which the option could be implemented
	3. Describing whether implementation of an option is underway and at what stage level
B. Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies. Lay language: how can implementation be improved?	1. Identifying and understanding barriers and facilitators to implement and option
	2. Identifying and understanding implementation strategies to deal or take advantage of barriers and facilitators
	3. Prioritizing barriers, facilitators, and implementation strategies

Technical language note: In implementation sciences, options (or interventions) can also be called **innovations** or **change management tools**. In implementation sciences, the implementation process could also be called **scale and spread.**

Implementing an option is a critical stage in the decision-making process. However, there are some interventions in which the implementation stage might not necessarily be critical (e.g., prescribing a clinical treatment course for a given hospitalized patient).

The conditions that an option requires to be implemented can be classified using behavioral (e.g., what individuals need to do for the option to be implemented) and/or contextual (that are often

split in relevant to the inner and outer settings) variables. The contextual variables, and the setting (i.e., inner and/or outer setting), include the potential equity implications that the implementation of a given option might have.

Stage 4. Monitoring implementation and evaluating impacts

This stage aims to monitor the implementation of a given option and to evaluate its causal impacts in a particular setting. It is structured as two different goals (A and B). This stage includes 7 different types of questions that may need to be answered (Table 5). Monitoring implementation and evaluating impacts can be done at the short, medium and/or long-term; identifying measurement strategies for problems and options are also a key part of this stage.

Table 2.5. Goals and types of questions for decision-making stage 4: monitoring implementation and evaluating impacts.

Goals	Types of question
A. Identifying measurement strategies for populations and outcomes. Lay language: how can we measure populations and results?	1. Identifying instruments to identify or categorize populations
	2. Choosing the most accurate instruments to identify or categorize populations
	3. Identifying measurement instruments for outcomes of interest
	4. Determining the best instruments to measure outcomes of interest
B. Monitoring and evaluating populations and outcomes of interests. Lay language: is it doing what is supposed to be doing?	1. Monitoring the implementation of an option or implementation strategy
	2. Evaluating the impact of an option or implementation strategy
	3. Interpreting the findings of monitoring implementation or evaluating the impact of an option or implementation strategy

Technical language note: Several frameworks build on evidence coming from this type of question to better understand the impact of a given intervention (e.g., **theory of change, logical framework**, etc.) and its mechanism of action.

2.7.Discussion

Principal findings and findings in relation to the existing literature

This paper develops a taxonomy of mutually exclusive and collectively exhaustive types of demand-driven questions in which evidence may provide decision-relevant insight. We identified forty different types of questions, which were classified across 14 different goals in four different decision-making stages. Some existing frameworks have been developed to formulate research questions, such as PICO (Santos et al., 2007) and SPIDER (Cooke et al., 2012), or to understand what type of categories of research questions can be addressed by evidence syntheses (Hunt et al., 2018). However, these frameworks were not built with a demand-driven approach complemented by existing frameworks to facilitate decision-making.

Although the field of knowledge translation has substantially evolved in recent decades, knowledge translation efforts and tools have concentrated on how new research findings can be better disseminated to decision-makers (Bhawra & Skinner, 2020). However, no available tools facilitate the interaction between decision-makers and evidence producers or intermediaries (i.e., people working in between researchers and decision-makers) at the question-formulation stage to achieve a more responsive evidence-support system.

A recently renewed focus on the co-production of knowledge – understood as a collaboration between evidence producers, decision makers, and any other stakeholder to design, implement and interpret research for a given need (Redman et al., 2021) – has of course yielded outputs that can support the future flow of new research. This taxonomy provides a more actionable output,

that could be used to help in co-produce evidence support. Hence, when a decision-making need emerges, collaborative work among decision-makers, evidence intermediaries and evidence producers facilitated by the taxonomy created in this paper might make easier to clarify the specific question for which an evidence-informed answer is required.

Strengths and limitations

This study has several strengths. First, this is the first paper that creates a mutually exclusive and collectively exhaustive list of types of question for which evidence could provide decision-relevant support. Secondly, the taxonomy was created using a demand-driven perspective by asking evidence-support groups to itemize the questions they have received from decision-makers. Hence, it is built from existing questions that have been addressed by at least one of a variety of operating evidence-support units. Finally, it uses generic language that facilitates the communication across different sectors/disciplines and different forms of evidence.

This study has also some limitations. First, it was infeasible to reach all the units that provide some type of support across all sectors and disciplines, and participants working in non-surveyed sectors might provide extensions to this taxonomy. Also, while this paper presents a mutually exclusive and collectively exhaustive list of types of question, it has not yet been applied to a specific setting or context to validate and facilitate the understanding of this taxonomy.

Implications for policy and practice

This taxonomy can have different implications depending on three main audiences. First, decision-makers (including government policymakers, professionals, and citizens) could easily scan the different types of questions to clarify the type of questions for which evidence could provide decision support. Second, impact-oriented evidence producers of any form of evidence could better orient their work to organize and prioritize types of questions, enhancing coordination and avoiding duplication among them. Finally, this tool could strongly support evidence-intermediaries in connecting the demand needs with the supply side.

When using this taxonomy of types of question, users should bear in mind the following considerations. First, although we have presented the types of question in a logical order, they are by no means intended as a list each of which those making policy decisions should consider for each one of their issues. Indeed, decision-makers can use one, some, or all of the questions to address a given issue. By providing guidance on what questions from this taxonomy would most usefully be addressed to answer a specific decision, evidence intermediaries could facilitate this selection.

Secondly, some types of question included might not be relevant for certain groups (e.g., comparing the importance of a problem against others in social sciences, or prioritizing spillover effects across different sectors). Thirdly, our aim in developing the taxonomy was to organize questions and not the results that research answering these questions could have. Hence, since they are essentially an assessment of the answer of a specific type of question, we considered

questions such as “What are the evidence gaps or the methodological limitations of the existing evidence for a given topic?” out of the scope. Finally, there are several types of question that are addressed by building on other complex frameworks (e.g., agenda setting of a policy issue (Kingdon, 2011); chances of a policy to be developed looking at institutions, interests, and ideas (National Collaboration Centre for Health Public Policy, 2014) or the political economy; or the external validity of a given body of evidence). These questions are important, and several types of questions from the taxonomy could contribute to conducting an assessment in these complex frameworks.

Implications for future research

This taxonomy of research questions is only a first of many efforts that could facilitate the connection between demand-side needs and evidence production and support. Further research should explore how different study designs could properly answer each type of question identified in this taxonomy. A concrete application of this taxonomy in a case study would help to validate and test the tool. Matching types of decisions (e.g., funding a new technology, what intervention to use for addressing a specific problem, whether acting now is the right time, conducting or not a pilot for a new technology) with the types of questions included in this taxonomy would, by specifying what types of question in this taxonomy should be answered depending on the specific type of decision, facilitate a stronger and more integrated evidence-support system.

Future research efforts could also involve going back to the survey participants and interview: (1) a sample of them to ask whether they have encountered additional questions that were not represented in the taxonomy, because they have been addressed by complementary groups in other sectors, or in groups that provide a more integrated evidence-support to decision-makers in a given country; and (2) other actors (e.g., government policymakers, science advisors, subject-matter experts, etc.) who could provide additional types of question that were not necessarily addressed by evidence advice.

Finally, future uses of the taxonomy in combination with artificial intelligence could consider these types of questions in their algorithms and quickly identify claims that are, or are not, supported by evidence.

2.8.Conclusion

This paper provides a unique taxonomy of 40 demand-driven types of questions where evidence could provide decision-relevant insights, structured around four decision-making stages (clarifying a societal problem, its causes, and potential impacts; finding and selecting options to address a problem; implementing or scaling-up an option; and monitoring implementation and evaluating impacts). Decision-makers, evidence intermediaries, and impact-oriented evidence producers could importantly benefit from this taxonomy to facilitate the exchange of evidence needs from decision-makers, through evidence intermediaries and to better connect evidence-production efforts among evidence producers.

2.9.Declarations

Ethics approval and consent to participate.

This study was approved by the Hamilton Integrated Research Ethics Board (HiREB), Project ID: 8279.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

The authors are grateful for the support received from the Global Commission on Evidence to Address Societal Challenges to conduct this work.

Authors contributions

CM and JNL discussed and created the idea. CM, AS, GG and JNL wrote the protocol. CM and JNL participated in the data collection, and everyone participated in the iterative data analysis. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge and thank all the individuals that participated giving their valuable feedback to make this taxonomy a list that would be useful to facilitate the better connection of research with decision making. They include Ludovic Reveiz, Michelle Haby, Kerry Albright, Tanja Kuchenmuller, Erik von Uexkull, Jeremy Grimshaw, Cristian Herrera, and the participants of the HSR 2022 Symposium, particularly Nasreen Jessani, Yodi Mahendradhata, Arash Rashidian, and Simon Lewin. We also thank Jennifer Verma, for her contributions in better formatting the figures of this paper.

2.10. References

Alonso-Coello, P., Schünemann, H. J., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., Treweek, S., Mustafa, R. A., Rada, G., Rosenbaum, S., Morelli, A., Guyatt, G. H., Oxman, A. D., & the GRADE Working Group. (2016). GRADE Evidence to Decision (EtD) frameworks: A systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*, i2016. https://doi.org/10.1136/bmj.i2016

Bhawra, J., & Skinner, K. (2020). Examination of tools associated with the evaluation of knowledge uptake and utilization: A scoping review. *Evaluation and Program Planning*, *83*, 101875. https://doi.org/10.1016/j.evalprogplan.2020.101875

Brownson, R. C., Royer, C., Ewing, R., & McBride, T. D. (2006). Researchers and Policymakers. *American Journal of Preventive Medicine*, *30*(2), 164–172. https://doi.org/10.1016/j.amepre.2005.10.004

Cochrane Convenes. (2022). *Preparing for and responding to global health emergencies: Learnings from the COVID-19 evidence response and recommendations for the future. February 2022*. https://doi.org/10.6084/m9.figshare.19115849.v2

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qualitative Health Research*, *22*(10), 1435–1443. https://doi.org/10.1177/1049732312452938

Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., & Lowery, J. (2022). The updated Consolidated Framework for Implementation Research based on user feedback. *Implementation Science*, *17*(1), 75. https://doi.org/10.1186/s13012-022-01245-0

Fretheim, A., Munabi-Babigumira, S., Oxman, A. D., Lavis, J. N., & Lewin, S. (2009). SUPPORT Tools for Evidence-informed Policymaking in health 6: Using research evidence to address how an option will be implemented. *Health Research Policy and Systems*, *7*(S1), S6. https://doi.org/10.1186/1478-4505-7-S1-S6

Global Commission on Evidence to Address Societal Challenges. (2022). *The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers*. McMaster Health Forum. https://www.mcmasterforum.org/docs/default-source/evidence-commission/evidence-commission-report.pdf?Status=Master&sfvrsn=2fb92517_5/Evidence-Commission-report

Global Commission on Evidence to Address Societal Challenges. (2022b). *The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers*. https://www.mcmasterforum.org/docs/default-source/evidence-commission/evidence-commission-report.pdf?Status=Master&sfvrsn=2fb92517_5/Evidence-Commission-report

Global Commission on Evidence to Address Societal Challenges. (2023). *Evidence Commission update 2023: Strengthening domestic evidence-support systems, enhancing the global evidence architecture, and putting evidence at the centre of everyday life*. McMaster Health Forum. https://www.mcmasterforum.org/docs/default-source/evidence-commission/update-2023.pdf?sfvrsn=e81cbf_8

Hunt, H., Pollock, A., Campbell, P., Estcourt, L., & Brunton, G. (2018). An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. *Systematic Reviews*, *7*(1), 39. https://doi.org/10.1186/s13643-018-0695-8

Innvær, S., Vist, G., Trommald, M., & Oxman, A. (2002). Health policy-makers' perceptions of their use of evidence: A systematic review. *Journal of Health Services Research & Policy*, *7*(4), 239–244. https://doi.org/10.1258/135581902320432778

Kingdon, J. W. (2011). *Agendas, Alternatives, and Public Policies* (Longman Classics Edition, Ed.). HarperCollins College Publishers.

Kuchenmüller, T., Lavis, J., Kheirandish, M., Reveiz, L., Reinap, M., Okeibunor, J., Siswanto, S., Rashidian, A., Sieber, S., Moat, K., Mansilla, C., El-Jardali, F., Helble, M., Reeder, J., Chapman, E., Barreto, J. O. M., Mandil, A., & Swaminathan, S. (2022). Time for a new global roadmap for

supporting evidence into action. *PLOS Global Public Health, 2*(6), e0000677. https://doi.org/10.1371/journal.pgph.0000677

Lavis, J. N., Wilson, M. G., Oxman, A. D., Grimshaw, J., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 5: Using research evidence to frame options to address a problem. *Health Research Policy and Systems*, *7*(Suppl 1), S5. https://doi.org/10.1186/1478-4505-7-S1-S5

Lavis, J. N., Wilson, M. G., Oxman, A. D., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 4: Using research evidence to clarify a problem. *Health Research Policy and Systems*, *7*(Suppl 1), S4. https://doi.org/10.1186/1478-4505-7-S1-S4

National Collaboration Centre for Health Public Policy. (2014). *Understanding Policy Developments and Choices Through the "3-i" Framework: Interests, Ideas and Institutions*.

Santos, C. M. da C., Pimenta, C. A. de M., & Nobre, M. R. C. (2007). The PICO strategy for the research question construction and evidence search. *Revista Latino-Americana de Enfermagem*, *15*(3), 508–511. https://doi.org/10.1590/S0104-11692007000300023

2.11. Additional files

Additional file 1. Questionnaire used to collect data from participants.

Identifying and categorizing health-policy research questions
Online survey
Questionnaire to be sent to units supporting the use of research evidence in decision making.

#	Question	Question type	Possible answers
1	Is your unit active in answering policymaking needs with research evidence?	Yes/no question	Yes or no
2	Please select the category that best describes the institution in which your unit is embedded.	Answer one option	National ministry Sub-national (e.g., provincial) ministry Government agency Legislative branch Judicial branch Non-governmental organization University Other (please specify)
3	Who is eligible to request evidence support to your unit?	Answer one or multiple options from a list	High-level policymakers (e.g., ministers, vice ministers) Mid-range policymakers (e.g., heads of units, departments) Managers of government agencies Staff in charge of program implementation People that are part of universities People working in NGOs. Other (Please specify)
4	Who most commonly request evidence support to your unit?	Answer one or multiple options from a list	High-level policymakers (e.g., ministers, vice ministers) Mid-range policymakers (e.g., heads of units, departments) Managers of government agencies Staff in charge of program implementation People that are part of universities People working in NGOs. Other (Please specify)
5	If you work in the health sector, which area(s) does your unit provide evidence support?	Answer one or multiple options from a list	Clinical practice Public health [including epidemiology] Technology assessments

#	Question	Question type	Possible answers
			Health systems Other (please specify)
6	If available, please provide the full list of research questions that your unit has addressed.	Free text	Free text
7	If possible, please provide the full list (with hyperlinks if available) of publications or reports produced to inform policymaking processes.	Free text	Free text

Additional file 2. Details of the types of question included in the taxonomy.

STAGE I. CLARIFYING A SOCIETAL PROBLEM, ITS CAUSES, AND POTENTIAL IMPACTS

This stage aims to clarify a given problem, identify potential causes, and outline potential impacts or spillover effects of it. It is structured by six different goals that may need to be achieved (A to F). In total, 15 different types of questions that may need to be answered to achieve the goal are included in this stage.

I.A. Choosing and prioritizing measurements of a problem (i.e., how can a problem be measured?)

This goal aims to get insights on what potential outcomes exist to characterize or measure a problem, the different values that individuals could have regarding the outcomes, and to prioritize what are the most suitable outcomes to characterize or measure the problem. Three types of questions are included in this goal:

I.A.1. Identifying measurements to characterize a problem.

Identifying measurements is a critical step of clarifying a problem. This type of question looks for different outcomes that are available and suitable to characterize or measure a given problem.

I.A.2. Understanding individuals' values regarding outcomes

People could have different values on the uses of different outcomes to characterize or measure a problem. Also, different stakeholders could have different values regarding outcomes (e.g., citizens may value different the outcomes than government policymakers).

I.A.3. Prioritizing measurements to characterize a problem.

The prioritization of measurements to characterize or measure a problem can be done by looking at the identified outcomes and weighting the different values and preferences of individuals regarding them.

I.B. Describing a problem and its magnitude (i.e., what's the problem and how big it is?)

This goal aims to describe a problem in a given moment of time and to identify the population that is affected by it. Two types of questions are included in this goal, namely.

I.B.1. Describing a problem in a point in time

The description of a problem or objective in a given moment of time can be done depending on the type of variable that would need to be measured. Hence, variables can be depicted, among other forms, using the following forms:

Frequencies (e.g., what is the (cumulated or new) frequency of a given characteristic within a population group?)

Central tendencies (e.g., what is the mean of a given characteristic within a population group?)

Distributions (e.g., what is the variation range of a given characteristic within a population group?)

By describing a problem using any of the mentioned variables, the magnitude of the problem is also presented and assessed.

Note about jargon: In epidemiological research, describing a problem through frequencies is often called **prevalence** (for number of people living with certain characteristics in a given moment of time) or **incidence** (for new number of people that started to have certain characteristics during a certain time). In clinical research, the frequency of signs and symptoms of a given condition is also called the **clinical presentation** of a disease.

I.B.2. Clarifying and characterizing populations affected by a problem.

To clarify a problem, the population that is affected by it might not always be clear. This type of question aims to specify the population and specific sub-populations that are affected (or most affected) by a certain problem.

I.C. Understanding a problem (i.e., how, and why is a problem?)

This goal aims to interpret a given problem by critically and conceptually analyzing it and understanding how it is perceived by different stakeholders and the role that the context has in the specific problem. Three types of questions are included in this goal:

I.C.1. Finding conceptual approaches to understand a problem.

Depending on the complexity of a given problem, conceptual frameworks might be useful to understand a problem, which might lead to better ways on how to address it. This question includes finding and selecting frameworks to describe and conceptualize a complex problem or issue, and how the problem is created by interconnecting with other factors (e.g., behavioral, and contextual variables).

I.C.2. Understanding stakeholders' perceptions of a problem

Exploring the perceptions that different groups have of a problem is critical to understand the problem and its magnitude. There are many types of stakeholders that might have different perceptions of the same issue, including government policymakers, organizational leaders, professionals, and citizens.

The interpretations of the results of a study answering this type of question can be used to understand the different framing of a given problem by a certain interest group.

Note about jargon: An important economic branch studies the role of **stated and revealed preferences** to distinguish between stakeholders' perceptions that are communicated ('stated') versus the ones that are not evidently communicated ('revealed').

I.C.3. Understanding the role of context in a problem

This question includes the understanding of the role that the context or the specific setting has in the problem. Contexts and/or settings can bring unique challenges/opportunities for a certain problem or objective to be developed or achieved.

I.D. Assessing the variability of a problem (i.e., how the problem varies over time, across populations and other problems?)

This goal aims to assess the variability or magnitude of a given problem comparing against time, other populations or settings, and other problems. The variability can be assessed by comparing only one variable (i.e., either time, other populations, or other problems) or multivariable (e.g., against time and other populations at the simultaneously). Three types of questions are included in this goal:

I.D.1. Assessing variability over time

Assessing the variability of a problem or issue when compared against time could be done over limited (e.g., 2 points) or multiple (e.g., time series) point of time. This type of question measures the evolution of an indicator or measurement over time.

I.D.2. Assessing variability across populations or settings

Assessing the variability of a problem or issue when compared against other populations or settings aims to explore how the problem is in other populations or settings. This type of question includes to identify what populations are most affected by the problem, or what populations are most at risk, which entails equity considerations.

I.D.3. Assessing the importance of a problem relative to other problems

Assessing the variability of a problem when compared against other problems entail to measure the relative importance (i.e., weight) of an issue compared to others.

Note about jargon: In epidemiological research, measuring the burden that a disease can cause compared to other is often called the **burden of disease**.

I.E. Understanding the causes and aggravating factors of a problem (i.e., what is causing or making the problem worse?)

In this goal, the problem is set as the dependant variable (being caused by something else) as opposed to goal F, where the problem is set as the independent variable (the problem being the cause of something else). However, in many cases, the framing of the question could be unclear (i.e., whether an issue is the problem or the cause of a different problem).

Note about jargon: In epidemiological research, causes can also be referred as **risk or protective factors** that individuals can experience when they are **exposed** to a certain cause. In public health research, some potential factors that could explain different health outcomes are called **determinants of health**. In clinical research, the factors that could explain better or worse clinical outcomes on a given health condition are commonly called **prognostic factors**, while the potential causes of a health condition are called the **etiology of the disease**. In some social sciences field, they could also be understood as **explanatory factors**, to understand what social factors would cause a given social behaviour.

This goal aims to identify and understand the relative importance of causes and/or aggravating factors of the problem. Two types of questions are included in this goal:

I.E.,1. Identifying causes and/or aggravating factors of a problem.

This question asks whether and to what extent a certain variable can be identified as a cause or aggravating factor of a certain problem. Identifying potential causes or aggravating factors of a problem require assessing whether a factor has potentially some effect on a certain outcome at different complexity levels. Different type of causes or aggravating factors might have variable effects on certain outcomes, and the level of exposure to a certain factor might also vary the effects on these outcomes. Finally, causes and aggravating factors might be of several types such as personal characteristics (e.g., age), contexts (e.g., living in rural areas), interventions received (e.g., mechanical ventilation), or

I.E.,2. Understanding the relative importance of causes and/or aggravating factors across population groups and contexts.

After having identified potential causes or aggravating factors of a problem, this type of question assesses the relative importance that causes and/or aggravating factors have based on their contribution to the problem, and to identify in what population groups the association between these causes and/or aggravating factors and the outcome is stronger.

I.F. Understanding the impacts of a problem (i.e., what impacts is the problem creating?)

As mentioned in goal I.C, in this goal the problem is set as an independent variable (the problem being the cause of something else), as opposed to goal I.F, where the problem is set as the dependent variable (the problem being caused by something else). A problem can create several types of impacts or spillover effects, that could be felt in one or multiple sectors (e.g., one health problem could create spillover effects on the environment).

Note about jargon: In economics, the unintended impacts of a given action are called **externalities** (e.g., passive smoking).

This goal aims to identify and prioritize the most important impacts or spillover effects of a given problem. Two types of questions are included in this goal:

I.F.1. Identifying impacts/spillover effects of a problem

This question asks whether and to what extent a given scenario is a consequence of a certain problem. Identifying potential impacts or spillover effects require assessing whether the problem has potentially some effect on a certain outcome at different complexity levels.

I.F.2. Prioritizing the most important impacts/spillover effects of a problem

After having identified potential impacts or spillover effects of a problem, this type of question assesses what are the most important of them, and to identify in what population groups would be most affected by these impacts.

STAGE 2. FINDING AND SELECTING OPTIONS TO ADDRESS A PROBLEM

This stage aims to find and select options that could address (or help to reduce the impact) of a problem. It is structured in four different goals that may need to be achieved (A to D). In total, 13 different types of questions that may need to be answered to achieve the goal are included in this stage.

II.A. Finding ad understanding potential options (i.e., what are the potential solutions?)

This goal aims to first identify a list of potential options to address a given problem, followed by understanding how and why they work. Two types of questions are included in this goal:

II.A.1. Scoping a list of potential options

A list of potentially available interventions to address a given problem should be the main output of answering this question. This type of question commonly includes a judgement of what options are potentially suitable for a specific contextual reality and asking what others (including what other jurisdictions) are doing to deal with a given problem.

Interventions might also not be available right now, but they might become available in future (e.g., vaccines under phase I-II trials).

Note about jargon: In clinical research, these potential options would be called **treatment alternatives** for treating a given health condition.

II.A.2. Understanding the way potential options and their components work

This type of question describes what does entail to develop a given option, including its mechanism of action (or causal pathway if applicable), and how and why it should work to address the problem or issue.

An important distinction should be made with an option that has been already implemented, and someone may want to know why it has not had the results that should have had. This issue will be re-taken in the stage 4 of this taxonomy (Monitoring implementation and evaluating impact).

II.B. Assessing the expected impact of options (i.e., is it feasible, should it work, is it convenient, and is it equitable and acceptable?)

This goal aims to assess the possible impact or success of options by assessing it in different outcomes. The impact of a given option can be assessed at one point in time, and can be assessed in the short, medium, and long-term to better assess the sustainability of the option.

This goal is sitting in the context in which an option has not yet been implemented, so we are assessing the possible impact of options. We will come back to the actual impact of the implementation of an option in stage 4 (Monitoring implementation and evaluating impact).

Also, the impact of an option can be assessed in different populations, so to well-formulate a question in this stage, someone should have a clear population in which the option is planning to be implemented.

Seven types of questions are included in this goal:

II.B.1. Assessing the feasibility of an option

Assessing this question entails making a judgement whether a given option is it going to be feasible to be implemented in a given context or setting, and it could be split in different dimensions (e.g., operationally feasible, legally feasible, etc.).

II.B.2. Assessing the benefits and early and frequently occurring harms of an option

An option can have benefits to address a problem or its causes. Assessing the benefits entails understanding the outcomes that need to be used to measure benefits, the effect size, and its variability.

Note about jargon: In epidemiological research, the benefits would be called **efficacy** (if it is measured under controlled circumstances) or **effectiveness** (if it is measured under 'real-world' circumstances).

II.B.3. Identifying late-occurring harms of an option

An option can also potentially have harms. Identifying them and assessing its probability of occurrence (i.e., risk) are a critical part of assessing the expected impact of options. In case the benefits of an option are also available, this type of question might entail assessing whether the benefits outweigh the harms (i.e., net benefit).

Harms can also be direct from the implementation of a given option or could also be unintentional or be more spillover effects of the implementation of a given option in other sectors or contexts.

Note about jargon: In health economics and health technology assessment, the outcome of assessing whether the benefits of a technology outweigh the harms is called **benefit-risk balance**.

II.B.4. Assessing the acceptability of an option

This type of question measures the level of acceptability that a given option would have in a concrete setting, and to what extent a given group is willing to receive an intervention.

Note about jargon: In some research disciplines, individuals' **satisfaction** with an option could also be included in this type of question.

II.B.5. Assessing the costs and resource use of an option

The potential costs and resource use that implementing a given option will create depend on the specific context and setting in which the option is planned to be implemented. Resources could be monetary (i.e., costs), but could also be human resources, technology, etc.

Note about jargon: In health economics and health technology assessment, calculating the costs that implementing a new technology will bring to the system is called **budget impact analysis**.

II.B.6. Assessing the efficiency in the use of resources

Using resources to implement an option creates questions related to how efficient these investments are. In this question, we compare whether the costs of implementing an option might be worth it.

Note about jargon: In economics, this efficiency in the use of resources is often called **value for money**. If a previous investment is conducted, **return on investment (ROI)** could also be used. In health economics, the **cost-effectiveness,** and the **incremental cost-effectiveness (cost-utility) ratio** of an intervention is often used to measure the efficiency in the use of resources.

II.B.7. Identifying equity, ethical and human rights impact of an option

Implementing an option might have several equity, ethical and human rights implications. In one side, implementing an option might have differential impact (measured as any of the other outcomes included in this goal) in some population groups. In the other side, some options could have ethical, social, and human rights implications when implemented that might arise when finding and selecting options. This type of question also includes the social value that a given option might have.

II.C. Maximizing the expected impact of options (i.e., how can we ensure success with these solutions?)

This goal aims to maximize the expected impact by either adjusting some variables of interventions, or by focusing the option on certain population groups or settings. Two types of questions are included in this goal:

II.C.1. Adjusting options and enabling factors to maximize impact.

This type of question evaluates whether adjusting some variables (e.g., the deliverer, the intensity of the intervention, etc.) could modify the expected impact (measured as any of the questions provided in II.B) of an option.

Note about jargon: In epidemiological research, variables that modify the impact of an intervention are called **modifiers**.

II.C.2. Finding population groups, settings, and contexts to focusing options

This type of question explores what setting or socioecological contexts, and/or in what population groups the intervention would produce most impact (measured as any of the questions provided in II.B, which includes in what population the intervention would achieve most equitable results). This includes understanding why the impact of an option is different in one context compared to another.

Note about jargon: In social sciences, this type of question would also include the analysis of **positive deviance**, that aims to understand why a reduced number of cases are producing positive results when others are not.

II.D. Contributing to prioritize and select options (i.e., how to prioritize or combine solutions?)

This goal aims to produce insights to select the best combination of options to address the problem or causes, by creating packages or creating a ranking of options.

It is important to notice that selecting what options to pursue would be out of the scope of this list, since many other non-evidence related factors could be considered to make a decision on what to implement, but this goal concentrated on the insights that evidence could provide to these specific types of decisions.

Two types of questions are included in this goal. They are not mutually exclusive, and, in fact, the ranking created could be a ranking of packages.

II.D.1. Creating packages of options

This type of question finds the right combination of interventions that would produce the optimal balance between the expected impacts (using any or a combination of the impacts described in II.B). It can also be framed in a way to ask what are the packages or options that produce a minimal threshold of impact.

II.D.2. Creating ranking of options

This type of question creates a ranking of options (or packages of options) sorted by the expected impact that they would produce (measured as any or a combination of the impacts described in II.B).

This type of question includes inquiries that are looking for the most impactful (e.g., most effective, least harmful) intervention to address a given problem.

STAGE 3. IMPLEMENTING OR SCALING-UP AN OPTION

This stage aims to address issues related to the implementation of a given option. It is structured in two different goals that may need to be achieved (A and B). In total, 6 different types of questions that may need to be answered to achieve the goal are included in this stage.

Note about jargon: In implementation sciences, options (or interventions) can also be called **innovations** or **change management tools**.

III.A. Planning and describing the implementation of an option (i.e., can it be done and what needs to happen to implement?)

This goal aims to plan and describe the implementation of a given option by identifying who has to do what to implement an option, what role the context has in the implementation process and what is the implementation level of a given option.

This goal looks at variables and conditions required for a given option to be implemented. These can be structured in behavioral (e.g., what individuals need to do for the option to be implemented) and contextual variables (that are often split in inner and outer settings). This can also be interpreted as what conditions are needed for an option to be feasible to be implemented.

Assuming that one potential decision of a new option might be to conduct a pilot or a implementing an option at small-scale, this goal also aims to draw on lessons learned from the early implementation of a given option to plan the scale-up of it.

Note about jargon: In implementation sciences, the implementation process could also be called **scale and spread.**

Three types of questions are included in this goal:

III.A.1. Identifying who has to do what to implement an option.

This question is looking at finding variables related to stakeholders' behaviors (behavioral variables) that are required for an option to be implemented (i.e., what people need to do in order to implement an option).

In order to answer this question, the identification of key stakeholders are one step to be conducted before.

III.A.2. Identifying the context in which the option could be implemented.

This question is looking at finding variables related to the inner or outer setting or context in which an option is feasible to be implemented (including the political climate).

III.A.3. Describing whether implementation of an option is underway and at what stage level.

Here we assess the level of implementation of an option (or group of options) in a given moment of time (or over time) in a concrete setting. This could also include how many population groups have been reached by the option over time, and the historical development of a given option in a given jurisdiction.

III.B. Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies (i.e., how the implementation can be improved?)

This goal aims to find barriers and facilitators, and implementation strategies that could address and take advantage of them, respectively. Barriers and facilitators can come from different domains (inner and outer settings, individuals (e.g., skills and capacities), etc.), from different levels (e.g., government, policy, system, etc.), and across different type of actors (e.g., service providers, users or patients, organizations, etc.).

It is not always clear whether a given variable constitutes a barrier or facilitator, or to whether a given variable could be framed in a positive (facilitator) or negative (barrier) way. Then, we can also call barriers and facilitators as implementation considerations.

The implementation process entails identifying strategies to ensure a sustainable implementation of an option. This means to apply implementation strategies that can take the form of interventions to address or take advantage of barriers and facilitators, or they could also be mitigation measures for the potential risks of implementing an option (e.g., risk-control or risk-management strategies), in the short, medium, and long-term.

Three types of questions are included in this goal:

III.B.1. Identifying and understanding barriers and facilitators to implement and option.

Barriers are variables that could block or delay the implementation of a given option. Facilitators are variables that could make the implementation of a given option easier. This question includes the identification of them as well as understanding why they are barriers and why the barriers could interfere or facilitate the implementation of an option.

III.B.2. Identifying and understanding implementation strategies to deal or take advantage of barriers and facilitators.

Implementation strategies to take advantage of them could accelerate the implementation process. This question includes the identification of them as well as understanding why they could work to take advantage of these facilitators.

III.B.3. Prioritizing barriers, facilitators, and implementation strategies.

Once barriers and facilitators have been identified, the assessment of the importance of several barriers and/or facilitators in the implementation of a given intervention in a specific setting is a critical step. Barriers and facilitators could be prioritized in order to reduce the scope and focus an implementation plan. At the same time, implementation strategies (i.e., interventions that could address barriers and take advantage of facilitators) could also be prioritized.

STAGE 4. MONITORING IMPLEMENTATION AND EVALUATING IMPACT

This stage aims to monitor the implementation of a given option and to evaluate the impact of a given option in a concrete setting. It is structured in two different goals that may need to be achieved (A and B). In total, 7 different types of questions that may need to be answered to achieve the goal are included in this stage.

IV.A. Identifying measurement strategies for populations and outcomes (i.e., how can we measure populations and results?)

This goal aims to identify and select measurement strategies for ascertain the right population and accurately measure the outcomes of interest. This is related to finding instruments to facilitate measurement in order to monitor implementation and evaluate impact. In ascertaining populations and measuring outcomes, the questions are split in first identifying available instruments, followed by the selection of the most suitable instrument. Four types of questions are included in this goal:

IV.A.1. Identifying instruments to identify or categorize populations.

Several measurement instruments might exist to identify the right population to monitor the implementation and/or evaluate the impact of an intervention.

IV.A.2. Choosing the most accurate instruments to identify or categorize populations.

After having identified potential instruments to identify the right population, this question evaluates the accuracy to identify them.

Note about jargon: In clinical research, this is used in the context of the identification of the population with a certain health condition, and it is called **diagnostic accuracy**.

IV.A.3. Identifying measurement instruments for outcomes of interest

Outcomes can be used to measure a problem or an option and can be used to measure in one or multiple points in time. This type of questions identifies the instruments available to measure outcomes.

IV.A.4. Determining the best instruments to measure outcomes of interest.

This question identifies how accurate a given instrument (e.g., measurement or scale) is to measure a given outcome of interest. This can be particularly helpful where there are complex phenomena to measure and where scales acting as proxys might be needed.

Note about jargon: In epidemiological research, this question is often split in **reliability** (the consistency in which a measurement brings the same results) and **validity** (the level of bias that a measurement could have).

IV.B. Monitoring and evaluating populations and outcomes of interests (i.e., is it doing what is supposed to be doing?)

This goal aims to monitor and evaluate populations and options or implementation strategies, including its sustainability over time. Here, we focus on scenarios where an option has been already implemented (as opposed to the questions included in stage 2). Two types of questions are included in this goal:

IV.B.1. Monitoring the implementation of an option or implementation strategy

In this type of question, the implementation of an option is monitored to see what the progress is.

IV.B.2. Evaluating the impact of an option or implementation strategy

Measuring the impact could be conducted by using any of the outcomes outlined in the question 2.B. This includes the assessment of potential unanticipated harms (or spillover effects) of an option or implementation strategy. In this type of question, the impact that an option actually had is measured.

IV.B.3. Interpreting the findings of monitoring implementation or evaluating the impact of an option or implementation strategy

This question includes asking why the option or implementation strategy is having the results that is showing and comparing them with the expected impact. It also includes the understanding of the pathway (and intermediate outcomes) for a given option or implementation strategy to produce impact, and to draw lessons learned in this process.

Note about jargon: Several frameworks build on evidence coming from this type of question to better understand the impact of a given intervention (e.g., **theory of change, logical framework**, etc.) and its mechanism of action.

Chapter 3: Matching the right study design to decision-maker questions: results from a Delphi study

3.1.Preface

As research evidence can play a critical role in many decision-making stages, evidence-support systems need to find efficient ways to connect evidence demand and supply. Building on the taxonomy presented on chapter 2, this chapter matches study designs to every question that was part of the taxonomy of demand-driven questions. A two-round online Delphi approach was used to reach consensus among methodological experts about the most suitable study designs for 28 types of question.

Dr John Lavis and I were responsible for conceiving the idea, and I designed the protocol, collected, and analyzed data, and wrote the manuscript. Dr. Gordon Guyatt provided critical insights on the study design, particularly on the design of the data collection instruments. Dr Arthur Sweetman provided insights on the best ways to analyze ranking-type questions. All co-authors provided feedback on the draft manuscript and the paper was submitted to PloS Global Public Health on 2 August 2023.

Matching the right study design to decision-maker questions: results from a Delphi study

Cristián Mansilla[1,2*], Gordon Guyatt[3], Arthur Sweetman[4], John N Lavis[1,3]

[1] McMaster Health Forum, McMaster University, Hamilton, Ontario, Canada.

[2] Health Policy PhD program, McMaster University, Hamilton, Ontario, Canada.

[3] Department of Health Research Methods Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

[4] Department of Economics, Faculty of Social Sciences, McMaster University, Hamilton, Ontario, Canada.

*Corresponding author

Email: camansil@gmail.com (CM)

Short title: Matching the right study design to decision-maker questions.

3.2.Abstract

Research evidence can play an important role in each stage of decision-making, evidence-support systems play a key role in aligning the demand for and supply of evidence. This paper provides guidance on what type of study designs most suitably address questions asked by decision-makers.

This study used a two-round online Delphi approach, including methodological experts in different areas, disciplines, and geographic locations. Participants prioritized study designs for each of 40 different types of question, with a Kendall's W greater than 0.6 and reaching statistical significance ($p<0.05$) considered as a consensus. For each type of question, we sorted the final rankings based on their median ranks and interquartile ranges and listed the four study designs with the highest median ranks.

Participants provided 29 answers in the two rounds of the Delphi and reached a consensus for 28 (out of the 40) questions (eight in the first round and 20 in the second). Participants achieved a consensus for 8 of 15 questions in stage I (clarifying a societal problem, its causes, and potential impacts), 12 of 13 in stage II (finding options to address a problem) and four of six in each of stages III (implementing or scaling-up an option) and IV (monitoring implementation and evaluating impact).

This paper provides guidance on what study designs are more suitable to give insights on 28 different types of questions. Decision-makers, evidence intermediaries (, researchers and funders

can use this guidance to make better decisions on what type of study design to commission, use or fund when answering specific needs.

3.3.Introduction

There has been a growing recognition of the potential for evidence to support many aspects of decision-making processes. Evidence can play a significant role in clarifying problems and their causes, analyzing potential solutions, supporting implementation, and monitoring implementation, and evaluating impacts (Fretheim et al., 2009; Lavis, Wilson, Oxman, Grimshaw, et al., 2009; Lavis, Wilson, Oxman, Lewin, et al., 2009).

During the COVID-19 pandemic, evidence provided multiple insights critical to questions asked by decision-makers (Pearson, 2021). The lessons learned from this global crisis led several organizations to call for formalizing and strengthening evidence-support systems and the global evidence architecture (Cochrane Convenes, 2022; Global Commission on Evidence to Address Societal Challenges, 2022; Kuchenmüller et al., 2022b).

Robust evidence-support mechanisms require alignment between decision-makers' needs (evidence demand) and the work of evidence producers (evidence supply) and intermediaries (i.e., people working in between researchers and decision-makers) (evidence supply). Typically, a given need can best be met with a combination of different forms of evidence. When answering complex questions, decision-makers often require a combination of different forms or lines of evidence (Global Commission on Evidence to Address Societal Challenges, 2023). For example, a policymaker may use quantitative data from a randomized controlled trial to describe the likely benefits and harms of implementing a particular policy intervention and may also rely on qualitative data to gain a deeper understanding of how the public might perceive the intervention.

Following the policy cycle used in Update 2023 from the Global Commission on Evidence to Address Societal Challenges, previous research efforts generated a taxonomy of decision-makers' questions grouped into four decision-making stages (UNPUBLISHED DATA) (Global Commission on Evidence to Address Societal Challenges, 2023). This taxonomy can help decision-makers to frame their need for evidence in a form that an evidence intermediary or producer can answer with different forms of evidence.

After identifying a clear decision-making need in the form of a question, evidence intermediaries and producers can tailor their evidence support to address specific issues and share evidence that is more likely to be used in decision-making processes. The current study builds on the previously created taxonomy of decision-maker questions and matches questions to study designs. While some have suggested a single evidence hierarchy with randomized trials and evidence syntheses at the top of an evidence pyramid (Evans, 2003), our approach aligns with work that presents a more sophisticated approach that recognizes that the preferred study design will differ based on the specific type of question being asked (Agoritsas et al., 2015; Petticrew, 2003).

This paper provides guidance on what type of study designs can address each decision-maker question. Specifically, it:

1. lists specific study designs that would provide some insights to address different types of question;

2. provides a ranking of study designs that would be most suitable to answer these questions.

3.4.Materials and Methods

This is a Delphi study of methodological experts in different areas and disciplines.

Participants

Methodological experts – from the six WHO regions (Africa, Americas, Eastern Mediterranean, Europe, South East Asia, and the Western Pacific) representing the eight different forms of evidence that were identified by the Global Commission on Evidence to Address Societal Challenges (data analytics, modelling, evaluation, behavioural/implementation research, qualitative insights, evidence syntheses, technology assessment/cost-effectiveness analysis and guidelines, and experience in the health sector and in other sectors) were identified through global networks (e.g., EVIPNet, Cochrane and GIN). The study team obtained contact information through publicly available sources, and purposively sampled experts using the above criteria. Between April and May 2023, they were invited to participate in this online Delphi study.

Data collection

We used an online questionnaire to ask participants to rank the suitability of study designs to answer different types of question. The questionnaire provided 40 different types of questions (Supplementary File 1 provides the entire taxonomy of types of question used in this study).

In both Delphi rounds, participants considered a list of study designs potentially relevant to each type of question and prioritized the designs' suitability to answer the question. To facilitate

understanding, a brief explanation of each type of question and some examples were also provided. A glossary of study designs was also made available for participants, and they could declare that they did not have enough expertise to answer a given question. Additionally, synonyms of study designs used in different disciplines were included in the questionnaire's options to rank.

In the first round, participants prioritized at least four study designs and suggested any additional study designs they considered to be missing from the original list. They could choose to work through the complete list of types of question or, depending on their expertise in the eight forms of evidence mentioned above, work through a subset of questions.

Using a ranking-type Delphi process (Kobus & Westner, 2016; Strasser, 2019), we used the level of consensus reached in round one to prioritize the questions included in the round two. Questions that reached a consensus in the first round were not included in the second round. In the second round, participants either confirmed the previous rank order from round 1 or suggested an alternative ranking. All participants in the second round answered the complete list of question types.

Statistical analysis

Responses were analyzed in each round using median ranks and distributions (interquartile ranges). Additionally, in each round, Kendall's W (a statistical measure of consensus in ranking-

type surveys) was used to measure consensus [12–15]. Answers were considered if the participants ranked at least two options and did not declare that they had insufficient expertise.

The ranking given by each participant was classified ordinally (e.g., first priority received a 1, second priority received a 2, etc.) and, since not every participant ranked all the options (they were asked to prioritize at least four options), non-ranked values were imputed as the average value between the latest item ranked and the greatest possible rank for each answer (e.g., if one participant ranked four of the six available options, the two options that were not ranked both received a value of 5.5). From these rankings, medians and interquartile ranges were calculated for each study design, and the Kendall's W with ties (considering missing values as ties, as suggested by Kendall & Gibbons (Maurice Kendall, Jean Dickinson Gibbons, 1990)) was calculated, along with their statistical significance, using the Real Statistics Resource Pack for Microsoft Excel (Zaiontc, C, 2020).

The study designs were sorted for each type of question based on their median ranks (from smallest to largest). If more than one option had the same median rank, then the smallest number of the 25^{th} percentile, followed by the smallest number of the 75^{th} percentile, was used to sort these options.

One question was considered to have reached consensus when Kendall's W was greater than 0.6 and statistically significant (p-value < 0.05). The options ranked lower in the list with its IQR suggesting no change in their ranking position (e.g., in a question of 7 study design options, a

cross-sectional study had a median rank of 7, and a IQR from 7 to 7) were not included in the second round. Suggestions of additional study designs collected from the first round were, when appropriate, included in the study designs to be ranked in the second round.

In each question, a study design whose median rank was 1, and its interquartile range does not alter its ranking as first were identified and placed as gold standard. Similarly, a study design whose median rank was last, and its interquartile range does not alter its ranking as last was also identified.

The Hamilton Integrated Research Ethics Board (HiREB), Project ID: 14691 approved the study. Participants received an information sheet including the ethical considerations of being part of this study, and their response was considered as implicit consent.

3.5.Results

Seventy-three methodological experts were initially identified as potentially eligible to participate and, to balance the sampling across WHO regions and forms of evidence, an invitation was sent to 46 individuals in the first round. Two participants declined the invitation to participate in the study and were replaced with two from the original list. Twenty-one participated in round 1 (response rate 46%), while 8 participated in round 2 (response rate 18%). In total, 29 answers were received in both rounds of the Delphi process.

Table 1 presents the participants' characteristics, including their WHO region and the forms of evidence assigned to them for each Delphi round. In the first round, most of the participants were from the Americas and Europe, while no experts participated from South-East Asia, and only one from the Eastern Mediterranean. In the second round, most of the participants were based in the Americas, while no participants were from South-East Asia or the Eastern Mediterranean.

Regarding forms of evidence, the share of participants was similar in both rounds. More than 80% of participants declared themselves as experts in evidence syntheses in both rounds, and approximately 60% reported having expertise in guidelines. The number of participants that identified themselves as having expertise in data analytics, modelling or technology assessments was reduced in the second round. All forms of evidence were represented with at least one participant in both rounds.

Table 3.1. Characteristics of the participants included in the study per each round (n = 21 for round 1; n= 8 for round 2), ranked by frequency in round 1

	Round 1		Round 2	
	N	**%**	**N**	**%**
WHO region				
Americas	8	38%	4	50%
Europe	5	24%	1	13%
Africa	4	19%	1	13%
Western Pacific	3	14%	2	25%
Eastern Mediterranean	1	5%	0	0%
South-East Asia	0	0%	0	0%
Form of evidence of expertise*				
Evidence syntheses	17	81%	7	88%
Behavioral/implementation research	13	62%	4	50%
Guidelines	12	57%	5	63%
Evaluation	11	52%	1	13%
Data analytics	7	33%	1	13%
Modelling	6	29%	1	13%
Qualitative insights	6	29%	3	38%
Technology assessments/cost-effectiveness analyses	4	19%	1	13%

*Participants could have expertise in more than one form of evidence. This means that percentages for this category might sum more than 100%

In round 1, the number of answers received for each type of the 40 different questions varied from 3 to 10, while the response rate varied from 13% to 83%. In round 2, the number of answers received varied from 6 to 8, while the response rate was consistently around 15% per question.

Table 2 shows the specific response rate for each type of question and the values of the Kendall's We reached in the first or second round of the Delphi process, for the types of question in which

a consensus was reached. In 28 types of question (out of 40), there was some evidence of consensus among the answers (Kendall's $W > 0.6$ and its p-value < 0.05) reached in either round 1 or 2. Supplementary File 2 presents details of the Kendall's W for each round, and the rank correlation (calculated from the Kendall's W), and for participants per question (including questions in which a consensus was not reached).

From the original list of types of question, because their Kendall's W was greater than 0.6 and were statistically significant in the first round (i.e., there was consensus), 8 original questions were not moved to the second round. Because there was agreement that their ranking was last, a number of study designs were removed from the first round. One new study design suggested by a participant was added in the second round.

Table 3.2 Response rate and Kendall's W for the types of question in which consensus was reached.

Stage	Goal	Type of question	N round 1 (response rate)	N round 2 (response rate)	KW reached
I. Clarifying a problem	Choosing and prioritizing outcomes of a problem	Identifying outcomes to characterize a problem	7 (37%)	8 (18%)	0.78
		Prioritizing outcomes to characterize a problem	5 (26%)	8 (18%)	0.70
	Describing a problem and its magnitude	Describing a problem in a point in time	3 (13%)	NA	0.70
		Clarifying and characterizing populations affected by a problem	4 (17%)	NA	0.71
	Understanding a problem	Understanding stakeholders' perceptions of a problem	9 (41%)	8 (18%)	0.66
	Assessing the variability of a problem	Assessing variability over time	3 (14%)	7 (16%)	0.80
	Understanding the impacts of a problem	Identifying impacts/spillover effects of a problem	6 (22%)	NA	0.96
		Prioritizing the most important impacts/spillover effects of a problem	6 (22%)	NA	0.90
II. Finding options	Finding and understanding potential options.	Scoping a list of potential options	4 (27%)	NA	0.76
		Understanding the way potential options and their components work	4 (27%)	7 (16%)	0.66
	Assessing the expected impact of options	Assessing the feasibility of an option	10 (28%)	7 (16%)	0.68
		Assessing the benefits and early-and-frequently occurring harms of an option	10 (28%)	6 (13%)	0.94
		Identifying late-occurring harms and risks of an option	10 (28%)	7 (16%)	0.72
		Assessing the costs and resource use of an option	10 (28%)	7 (16%)	0.73
		Assessing the efficiency in the use of resources	7 (19%)	NA	0.80
		Identifying equity, ethical, social, and human rights impact of an option	10 (28%)	7 (16%)	0.79
		Assessing the acceptability of an option	10 (28%)	7 (16%)	0.81
	Maximizing the expected impact of options	Finding population groups, settings, and contexts to focusing options	7 (21%)	7 (16%)	0.62
		Creating packages of options	7 (25%)	7 (16%)	0.76

	Contributing to prioritize and select options	Creating a ranking of options	7 (25%)	NA	0.61
III. Implementing AN option	Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies	Identifying and understanding barriers and implementation strategies to deal with them	6 (38%)	7 (16%)	0.60
	Planning and describing the implementation of an option	Identifying who has to do what to implement an option	5 (83%)	7 (16%)	0.84
		Identifying the context in which the option could be implemented	4 (67%)	7 (16%)	0.75
		Describing whether implementation of an option is underway and at what stage level	4 (67%)	7 (16%)	0.72
IV. Monitoring and evaluation	Identifying measurement strategies for populations and outcomes	Identifying instruments to ascertain populations	4 (24%)	6 (13%)	0.67
		Identifying measurement instruments for outcomes of interest	4 (24%)	6 (13%)	0.71
	Monitoring and evaluating populations and outcomes of interests	Measuring the impact of an option or implementation strategy	6 (27%)	NA	0.63
		Interpreting the findings of measuring the impact of an option or implementation strategy	5 (23%)	6 (13%)	0.84

KW: Kendall's W; NA: Not available (question not included in the second round)

Tables 3 to 6 show the first four study designs in which a consensus was reached to address each one of the 28 types of question. Tables are structured depending on the specific decision-making stage in which they were classified in the original taxonomy of types of question. Supplementary file 3 shows the full ranking of study designs for all the 40 questions (including the questions in which a consensus was not reached).

In stage I (clarifying a societal problem, its causes, and potential impacts), shown in Table 3, 8 (out of 15) types of question reached a consensus. A clear gold standard (i.e., a study design the median rank of which was 1, and its interquartile range does not alter its ranking as first) was identified in five of the questions included, whereas in all of the types of question one or more study designs were clearly identified as a last priority (i.e., a study design whose median rank was last, and its interquartile range does not alter its ranking as last).

Stage I has most of the types of questions in which a consensus was not reached (7). These questions were related to understanding stakeholders' values regarding outcomes, describing a problem and its magnitude at a point in time, understanding the role of the context in a given problem, assessing in the variability of the problem across populations and locations and relative to other problems, and understanding the causes and/or aggravating factors of a problem.

Table 3.3 Ranking of the first four study designs to address the types of question in which consensus was reached in stage 1.

Goal	Type of question	Study design	Median rank (IQR)
Choosing and prioritizing outcomes of a problem	Identifying outcomes to characterize a problem	Review of outcomes that have been used by other studies (e.g., scoping review)	1 (1, 1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experience	2 (2, 4)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of experts' opinion	3 (3, 3)
	Prioritizing outcomes to characterize a problem	Delphi study (to get consensus from experts)	1 (1, 1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	2 (2, 2.25)
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions are using	5 (2.75, 6)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	3.5 (3, 5.25)
Describing a problem and its magnitude	Describing a problem during a period of time	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	1.5 (1, 2.25)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	2 (1.75, 2.25)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	3 (2.5, 3.25)
		Delphi study (to get consensus from experts)	6.5 (5.38, 6.63)
Understanding a problem	Finding conceptual approaches to understand a problem	Review to identify existing frameworks (conceptual analysis)	1 (1, 1)
		Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	2 (2, 3.25)
		Review to build a new framework (critical interpretive synthesis)	3 (2.75, 3)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	4 (4, 4)
	Understanding stakeholders' perceptions of a problem	Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	1 (1, 1.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	2 (2, 2)

		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	3 (3, 3.25)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4 (4, 4.25)
Assessing the variability of a problem	Assessing variability over time	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	1 (1, 1)
		Descriptive (not predicting) time-series analysis (including trend analysis)	2 (2, 2)
		Modelling to predict future scenarios (e.g., system dynamics, ARIMA models, etc.)	3 (3, 4.5)
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	4 (3.25, 4)
		Case reports (case series)	5 (5, 5.75)
Understanding the impacts of a problem	Identifying impacts/spillover effects of a problem	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3.5 (1.5, 7.75)
		Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	3.5 (3, 7.75)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	5 (1, 9)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	6.5 (2.5, 9)
	Prioritizing the most important impacts/spillover effects of a problem	Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	4.5 (3.25, 6.5)
		Delphi studies (to get consensus from experts)	5.5 (1, 10.38)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	7.25 (2, 9.88)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	7.75 (3, 9.88)

In stage II (finding options to address a problem), shown in Table 4, 12 (out of 13) types of question reached a consensus. A clear gold standard was clearly identified in 7 of the questions, whereas in only one of them was at least one study design not identified as clear last priority. There was only one type of question in which a consensus was not reached, which was related to adjusting options to maximize their impacts.

Table 3.4 Ranking of the first four study designs to address the types of question in which consensus was reached in stage 2.

Goal	Type of question	Study design	Median rank (IQR)
Finding and understanding potential options.	Scoping a list of potential options	Review to find options that have been used by other studies (e.g., scoping review)	1 (1, 1)
		Jurisdictional scan (comparative analysis) to understand what options have been implemented by other jurisdictions	2 (2, 2.25)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions	3.5 (2.75, 4.25)
		Ecological study (population-based study, including spatial analysis)	3.75 (3, 4.63)
	Understanding the way potential options and their components work	Randomized-controlled study (randomized experiment or randomized trial) measuring intermediate outcomes	1 (1, 3.5)
		Review to identify existing frameworks (conceptual analysis) that explain how an intervention might work	2 (1, 2)
		Interrupted time-series analysis (including joint-point regression) measuring intermediate outcomes	3 (3, 4.5)
		Descriptive (non-qualitative) case study	4 (4, 4)
Assessing the expected impact of options	Assessing the feasibility of an option	Delphi studies (to get consensus from experts)	1 (1, 2)
		Jurisdictional scan (comparative analysis) to understand the feasibility of the option elsewhere	2 (1, 2)
		Discrete choice experiment (stated preferences)	3 (3, 5.75)
		Modelling to predict whether the option will be feasible (e.g., system dynamics, ARIMA models, etc.)	4 (4, 6.75)
	Assessing the benefits and early-and-frequently occurring harms of an option	Randomized-controlled study (randomized experiment or randomized trial)	1 (1, 1)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	2 (2, 2)
		Interrupted time-series analysis (including joint-point regression)	3 (3, 3)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	4 (4, 5.5)
	Assessing late-occurring harms and risks of an option	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study), including databases of adverse event reporting (e.g., pharmacovigilance)	1 (1, 3.5)
		Randomized-controlled study (randomized experiment or randomized trial)	2 (1, 2)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	2 (2, 3)
		Case-control study (case-comparison study)	4 (4, 5.5)

	Assessing the costs and resource use of an option	Modelling to estimate the cost of an option	1 (1, 1)
		Jurisdictional scan (comparative analysis) to understand the costs in other jurisdictions	2 (2, 3)
		Case reports (case series)	3 (2, 7)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	4 (4, 6)
	Assessing the efficiency in the use of resources	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses)	1 (1, 1)
		Jurisdictional scan (comparative analysis) to understand whether the option was efficient in other jurisdictions	2 (2, 2.5)
		Delphi studies (to get consensus from experts)	3 (2.5, 3)
	Identifying equity, ethical, social, and human rights impact of an option	Delphi studies (to get consensus from experts)	1 (1, 1.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	2 (2, 2)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	3 (2.5, 3)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4 (4, 4)
	Assessing the acceptability of an option	Discrete choice experiment (stated preferences)	1 (1, 1.5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	2 (2, 2)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	3 (3, 3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	4 (4, 4)
Maximizing the expected impact of options	Finding population groups, settings, and contexts to focusing options	Case-control study (case-comparison study)	1.5 (1, 2.75)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	2 (1.25, 2)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3 (3, 5.25)
		Randomized-controlled study (randomized experiment or randomized trial) using subgroup comparisons.	4 (2.5, 4)
Contributing to prioritize	Creating packages of options	Evidence synthesis of studies evaluating the impact of single interventions to analyze the combined effect of packages.	1 (1, 1)

and select options		Randomized-controlled study (randomized experiment or randomized trial) to compare packages of interventions in different arms	2 (2, 2)
		Randomized-controlled study (randomized experiment or randomized trial) using posthoc comparisons.	3 (3, 3)
		Delphi study (to get consensus from experts)	4 (4, 4)
	Creating a ranking of options	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses) to create a ranked list of options	1 (1, 2)
		Ranking type Delphi study (to get consensus from experts)	2 (2, 2.5)
		Multi-criteria (objective) decision analysis to create a ranked list of options	2 (2, 3)
		Discrete choice experiment (stated preferences)	4 (4, 4.75)

In stage III (implementing or scaling-up an option), four (out of six) types of questions reached a consensus. A clear gold standard was identified in all four questions included. In only one question there was not at least one study design identified as clear last priority. Two types of questions did not reach a consensus and were related to the identification and prioritization of facilitators and implementation strategies to take advantage of them.

Table 3.5 Ranking of the first four study designs to address the types of question in which consensus was reached in stage 3.

Goal	Type of question	Study design	Median rank (IQR)
Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies	Identifying and understanding barriers and implementation strategies to deal with them	Review to find barriers and implementation strategies that have been used by other studies (e.g., scoping review)	1 (1, 1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	2 (2, 2.5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4 (3.5, 4.75)
		Social network analysis (mapping network analysis) to identify barriers and implementation strategies	5 (3, 5.75)
Planning and describing the implementation of an option	Identifying who has to do what to implement an option	Delphi studies (to get consensus from experts)	1 (1, 1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	2 (2, 2)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	3 (3, 4.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4 (4, 4)
	Identifying the context in which the option could be implemented	Jurisdictional scan (comparative analysis) to understand what other jurisdictions have identified as contextual variables	1 (1, 1)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	2 (2, 3.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	3 (3, 3.5)
		Delphi studies (to get consensus from experts)	4 (3, 4)
	Describing whether implementation of an option is underway and at what stage level	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	1 (1, 1)
		Descriptive (not predicting) time-series analysis (including trend analysis)	2 (2, 3.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	3 (2.5, 3)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches, documentary review of public speeches, etc.)	4 (4, 4.75)

In stage IV (monitoring implementation and evaluating impacts), four (out of six) types of questions reached a consensus. A clear gold standard was identified in three of the questions, and in two of them at least one study design was identified as clear last priority.

There were two types of questions where a consensus could not be reached. These They were related to the first goal (identifying measurement strategies for populations and outcomes). While the two other questions in this goal reached a consensus, the questions related to choosing the most accurate approach, and determining the best instruments, to measure/ascertain populations and outcomes did not reach consensus.

Table 3.6 Ranking of the first four study designs to address the types of question in which consensus was reached in stage 4.

Goal	Type of question	Study design	Median rank (IQR)
Identifying measurement strategies for populations and outcomes	Identifying instruments to ascertain populations	Review to find measurement strategies that have been used by other studies (e.g., scoping review)	1 (1, 3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions on measurement strategies	2 (1, 3)
		Jurisdictional scan (comparative analysis) to understand what measurement strategies have been used by other jurisdictions	2 (2, 2)
		Descriptive (non-qualitative) case study	4 (4, 5)
	Identifying measurement instruments for outcomes of interest	Review to find measurement strategies that have been used by other studies (e.g., scoping review)	1 (1, 1)
		Jurisdictional scan (comparative analysis) to understand what measurement strategies have been used by other jurisdictions	2 (2, 2)
		Descriptive (non-qualitative) case study	3 (3, 3.75)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions on measurement strategies	4 (4, 4)
Monitoring and evaluating populations and outcomes of interests	Measuring the impact of an option or implementation strategy	Randomized-controlled study (randomized experiment or randomized trial), including pragmatic trials	1 (1, 1)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3.5 (2.25, 4)
		Interrupted time-series analysis (including joint-point regression)	3.5 (3, 4)
		Regression discontinuity study (regression kink study or analysis)	5 (2.5, 8.625)
	Interpreting the findings of measuring the impact of an option or implementation strategy	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	1 (1, 1)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	2 (2, 2)
		Descriptive (non-qualitative) case study	3 (3, 3)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4 (4, 4)

3.6.Discussion

Principal findings and findings in relation to the existing literature

This study used a two-round Delphi process to produce a list of preferred study designs that investigators can use to address 28 different types of questions that were part of a mutually exclusive and collectively exhaustive taxonomy of types of questions in which evidence could provide support (UNPUBLISHED DATA). To reach a consensus on preferred study designs, the study elicited opinions from 29 experts.

In stage 1 (clarifying a problem, its causes and potential impacts), five types of question had clear preferred study designs (scoping reviews to identifying outcomes; Delphi studies to prioritize outcomes; reviews of frameworks to find conceptual approaches; qualitative inductive designs to understand stakeholders' perceptions and retrospective cohort study to assess the variability of a problem over time, while seven types of question (scoping reviews to scope a list of potential options, randomized-controlled study to assess benefits and early-and-frequently occurring harms, modelling to assess the costs, economic evaluations to assess the efficiency in the use of resources, Delphi studies to identify equity, ethical and social and human rights impacts, discrete choice experiments to assess the acceptability, and evidence syntheses to create packages of options) fill this criteria in stage 2 (finding and selecting options). In stage 3 (implementing an option), four types of question had a preferred study design (scoping reviews to identify and understand barriers, Delphi studies to identify who has to do what, jurisdictional scans to identify the context in which the option could be implemented, and cross-sectional studies to describe whether implementation is underway), while three types of question (scoping reviews to identify measurement instruments, randomized-controlled studies to measure the

impact of an option, and cross-sectional study to interpret the findings of measuring the impact) fill this criteria in stage 4 (monitoring implementation and evaluating impacts).

A consensus was not reached in 12 of the 40 different types of questions, the majority of which were part of the first decision-making stage (clarifying a societal problem, its causes, and potential impacts). In the first round of the Delphi, 8 types of questions reached a consensus and were not included in the subsequent stage, while 20 reached a consensus in the second stage when they had not reached a consensus in the first stage.

Previous research efforts have generated hierarchies or guidance of evidence for some types of questions or study designs (Daly et al., 2007; Parkhurst & Abeysinghe, 2016) and includes a more pragmatic approach to build guidance on the suitability of study designs to address demand-driven types of question. In this sense, this study moves beyond a static hierarchy or guidance of study designs, assuming that different study designs might more suitably answer different types of question. Regardless of the acceptance of a hierarchy, typology or a list of study designs, it is largely accepted that the traditional hierarchy of evidence (having RCTs and systematic reviews on top of a pyramid) could only provide insights for one type of question (which is represented twice in this paper as “Assessing the benefits and early-and-frequently occurring harms of an option”, and “Measuring the impact of an option or implementation strategy”) and that different types of study designs would be more appropriate than others based on the specific type of question that is to be addressed (Evans, 2003; Petticrew, 2003).

Strengths and limitations

This paper has a number of strengths. First, it creates a list of preferred study designs that can answer specific types of questions collected using a demand-driven approach (i.e., hence, are more likely to be related to the questions that decision-makers might ask). Secondly, it does so in a way that gave voice to various methodological experts, sampling by eight different forms of evidence and WHO region. This included capturing the different names were used for the same study designs, and the same names used for different study designs, depending on the specific discipline, which is also reflected in this study. The glossary of study designs that was used as a complement to the questionnaire also included multiple synonyms of study designs that could speak to multiple forms of evidence. Finally, it also does so using a robust methodology to reach a strong consensus and identify critical areas in which further work is required to reach a consensus (if possible), using a two-round Delphi process conducted online to reach broader audiences.

This study also has limitations. First, Delphi studies have the potential of acquiescence bias, which is the tendency of the participants to agree with the existing result, and thus failing to address their true preference. Secondly, since we prioritized the questions, each expert received based on their methodological expertise, the methodological experts answered different sets of questions included in the first round of the Delphi. However, this contributed to reaching a higher response rate, particularly in round 1. Thirdly, despite all the efforts were made to have a higher response rate, round two of the Delphi reached a low response rate. Finally, while we used Kendall's W and rank correlation to quantify the consensus reached in ranking-type surveys, these statistical measures have potential biases, such as not properly discriminating the

distance between ranks (e.g., the disagreement arising from ranking in the sixth versus seventh place is not different from the disagreement arising from ranking in the first or second place).

Implications for policy and practice

Evidence intermediaries can use the results of this paper in their role of promoting the use of evidence in decision-making processes. By identifying the type of question that decision makers are asking, they can either search or commission a study using a design that would be well suited to address that particular need. Researchers can use the results of this paper to guide them in choosing the preferred study design for the type of question at hand, or the right combination of them (i.e., in a mixed-methods study) to address a specific decision-making need. Finally, funders can use this guidance to decide what type of study design to fund or use, depending on the specific need at hand.

Implications for future research

Future research could explore several different areas. Firstly, subsequent efforts can explore reaching a consensus on the 12 types of questions in which a consensus was not reached. While a third round of the Delphi process could have been conducted, we might have a very limited number of responses. Alternatively, we suggest conducting a series of structured meetings (that follows the approach used by the GRADE Working Group to reach a consensus (G. H. Guyatt et al., 2008)) in which methodological experts can discuss what approaches can be better for specific types of questions.

Secondly, future studies could explore how considerations beyond study design (e.g., data analysis for any given study design) could play a role in answering specific types of question, and particularly how methodological limitations, or credibility of the findings, could also intersect with the selection of a study design. Thirdly, further research could weigh the role of domestic versus global evidence, and primary studies versus evidence syntheses as complementary approaches to provide an evidence-informed answer for specific types of question. Similarly, future efforts could consider different forms of evidence that are not necessarily part of research efforts, and in which the selection of study design might not have the same relevance. Finally, while qualitative study designs were included as part of this study, they were only part of a high-level category based on their paradigm (deductive/inductive) and general aim (to describe/ to critically interpret). Future research could go beyond this and use specific qualitative study designs to understand better what type of study designs are better than others to address a given type of question.

3.7.References

Agoritsas, T., Vandvik, P. O., Neumann, I., Rochwerg, B., Jaeschke, R., Hayward, R., Guyatt, G., & McKibbon, K. A. (2015). Finding Current Best Evidence. In G. Guyatt, D. Rennie, M. O. Meade, & D. J. Cook (Eds.), *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed* (1–Book, Section). McGraw-Hill Education. jamaevidence.mhmedical.com/content.aspx?aid=1183875650

Cochrane Convenes. (2022). *Preparing for and responding to global health emergencies: Learnings from the COVID-19 evidence response and recommendations for the future. February 2022*. https://doi.org/10.6084/m9.figshare.19115849.v2

Daly, J., Willis, K., Small, R., Green, J., Welch, N., Kealy, M., & Hughes, E. (2007). A hierarchy of evidence for assessing qualitative health research. *Journal of Clinical Epidemiology*, *60*(1), 43–49. https://doi.org/10.1016/j.jclinepi.2006.03.014

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84. https://doi.org/10.1046/j.1365-2702.2003.00662.x

Fretheim, A., Munabi-Babigumira, S., Oxman, A. D., Lavis, J. N., & Lewin, S. (2009). SUPPORT Tools for Evidence-informed Policymaking in health 6: Using research evidence to address how an option will be implemented. *Health Research Policy and Systems*, *7*(S1), S6. https://doi.org/10.1186/1478-4505-7-S1-S6

Global Commission on Evidence to Address Societal Challenges. (2022). *The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers*. https://www.mcmasterforum.org/docs/default-source/evidence-commission/evidence-commission-report.pdf?Status=Master&sfvrsn=2fb92517_5/Evidence-Commission-report

Global Commission on Evidence to Address Societal Challenges. (2023). *Evidence Commission update 2023: Strengthening domestic evidence-support systems, enhancing the global evidence*

architecture, and putting evidence at the centre of everyday life. McMaster Health Forum. https://www.mcmasterforum.org/docs/default-source/evidence-commission/update-2023.pdf?sfvrsn=e81cbf_8

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Kobus, J., & Westner, M. (2016). *Ranking-type delphi studies in IS research: Step-by-step guide and analytical extension*.

Kuchenmüller, T., Lavis, J., Kheirandish, M., Reveiz, L., Reinap, M., Okeibunor, J., Siswanto, S., Rashidian, A., Sieber, S., Moat, K., Mansilla, C., El-Jardali, F., Helble, M., Reeder, J., Chapman, E., Barreto, J. O. M., Mandil, A., & Swaminathan, S. (2022). Time for a new global roadmap for supporting evidence into action. *PLOS Global Public Health*, *2*(6), e0000677. https://doi.org/10.1371/journal.pgph.0000677

Lavis, J. N., Wilson, M. G., Oxman, A. D., Grimshaw, J., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 5: Using research evidence to frame options to address a problem. *Health Research Policy and Systems*, *7*(Suppl 1), S5. https://doi.org/10.1186/1478-4505-7-S1-S5

Lavis, J. N., Wilson, M. G., Oxman, A. D., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 4: Using research evidence to clarify a problem. *Health Research Policy and Systems*, *7*(Suppl 1), S4. https://doi.org/10.1186/1478-4505-7-S1-S4

Maurice Kendall, Jean Dickinson Gibbons. (1990). *Rank correlation methods* (5th Edition). Oxford University Press.

Parkhurst, J. O., & Abeysinghe, S. (2016). What Constitutes "Good" Evidence for Public Health and Social Policy-making? From Hierarchies to Appropriateness. *Social Epistemology*, *30*(5–6), 665–679. https://doi.org/10.1080/02691728.2016.1172365

Pearson, H. (2021). How COVID broke the evidence pipeline. *Nature*, *593*(7858), 182–185. https://doi.org/10.1038/d41586-021-01246-x

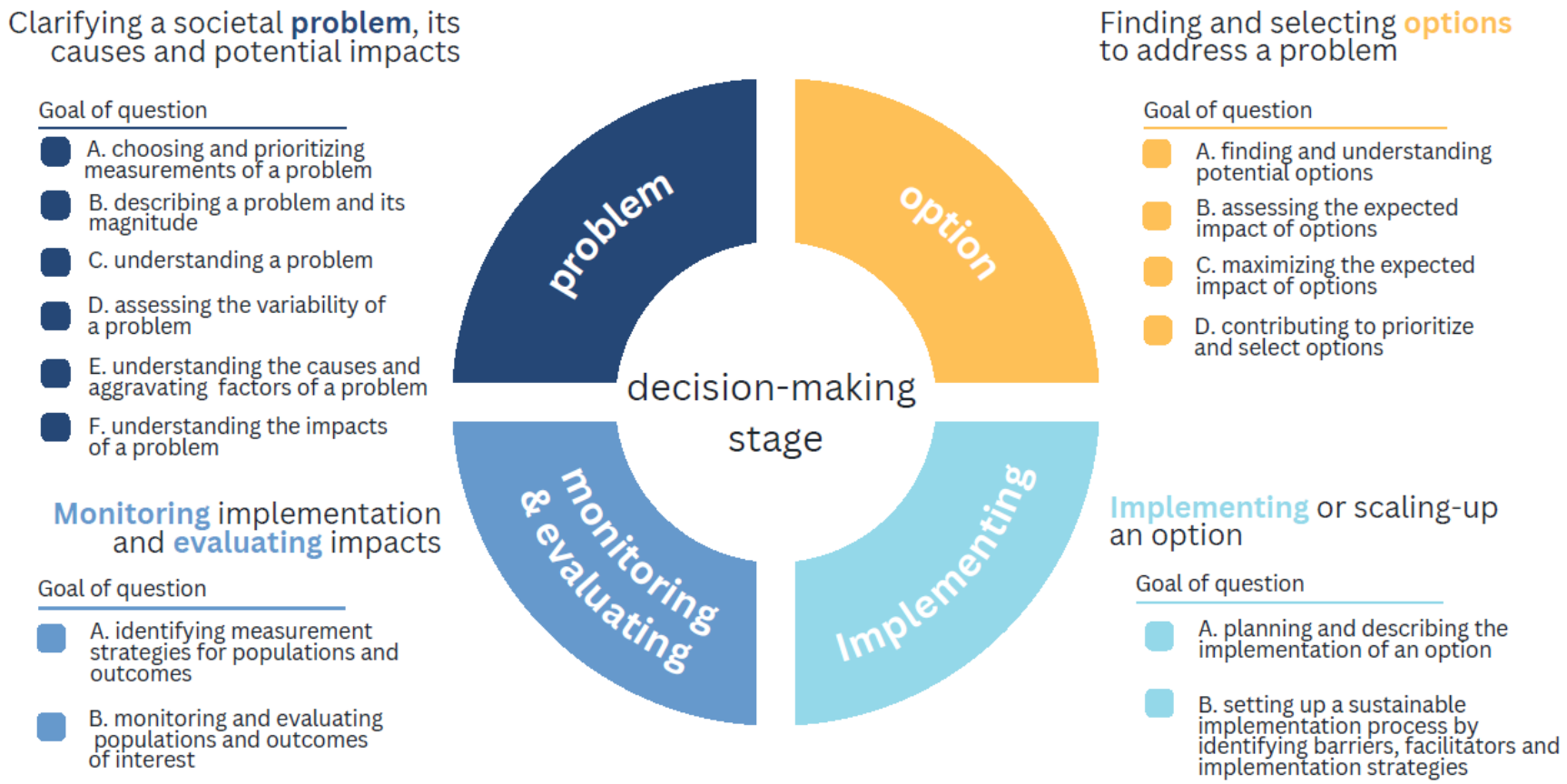
Petticrew, M. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527–529. https://doi.org/10.1136/jech.57.7.527

Strasser, A. (2019). Design and evaluation of ranking-type Delphi studies using best-worst-scaling. *Technology Analysis & Strategic Management*, *31*(4), 492–501. https://doi.org/10.1080/09537325.2018.1521956

Zaiontc, C. (2020). *Real Statistics Using Excel* [Computer software]. www.real-statistics.com

3.8.Supporting Information

S1 Figure. A demand-driven taxonomy of the types of question that could be addressed by evidence.



S2 table. Response rates and consensus reached in each Delphi round for each question.

Stage	Goal	Type of question	Round 1			Round 2			
			N (response rate)	KW**	RC***	N (response rate)	% of agree with last rank	KW**	RC***
I. Clarifying a problem	A. Choosing and prioritizing outcomes of a problem	1. Identifying outcomes to characterize a problem	7 (36.8%)	0.30	0.18	8 (17.8%)	63%	0.78*	0.75
		2. Understanding individuals' values regarding outcomes	6 (31.6%)	0.16	-0.01	8 (17.8%)	38%	0.45*	0.37
		3. Prioritizing outcomes to characterize a problem	5 (26.3%)	NA	NA	8 (17.8%)	50%	0.70*	0.65
	B. Describing a problem and its magnitude	1. Describing a problem in a point in time	3 (12.5%)	0.70*	0.55				
		2. Clarifying and characterizing populations affected by a problem	4 (16.7%)	0.71*	0.62				
	C. Understanding a problem	1. Finding conceptual approaches to understand a problem	10 (45.5%)	0.43*	0.36	8 (17.8%)	63%	0.53*	0.46
		2. Understanding stakeholders' perceptions of a problem	9 (40.9%)	0.16	0.05	8 (17.8%)	63%	0.66*	0.61
		3. Understanding the role of context in a problem	9 (40.9%)	0.20	0.10	8 (17.8%)	38%	0.29*	0.18
	D. Assessing the variability of a problem	1. Assessing variability over time	3 (14.3%)	0.17	-0.25	7 (15.6%)	43%	0.80*	0.77
		2. Assessing variability across populations and locations	4 (19%)	NA	NA	7 (15.6%)	43%	0.49*	0.40
		3. Assessing the importance of a problem relative to other problems	4 (19%)	0.43	0.24	7 (15.6%)	43%	0.52*	0.44

	E. Understanding the causes and aggravating factors of a problem	1. Identifying causes and/or aggravating factors of a problem	5 (22.7%)	NA	NA	7 (15.6%)	29%	0.52*	0.45
		2. Understanding the relative importance of causes and/or aggravating factors across population groups	5 (22.7%)	NA	NA	6 (13.3%)	33%	0.45*	0.34
	F. Understanding the impacts of a problem	1. Identifying impacts/spillover effects of a problem	6 (22.2%)	0.96*	0.96				
		2. Prioritizing the most important impacts/spillover effects of a problem	6 (22.2%)	0.90*	0.88				
II. Finding options	A. Finding and understanding potential options.	1. Scoping a list of potential options	4 (26.7%)	0.76*	0.68				
		2. Understanding the way potential options and their components work	4 (26.7%)	NA	NA	7 (15.6%)	57%	0.66*	0.61
	B. Assessing the expected impact of options	1. Assessing the feasibility of an option	10 (27.8%)	0.08	-0.02	7 (15.6%)	43%	0.68*	0.62
		2. Assessing the benefits and early-and-frequently occurring harms of an option	10 (27.8%)	0.26*	0.18	6 (13.3%)	50%	0.94*	0.93
		3. Identifying late-occurring harms and risks of an option	10 (27.8%)	0.13	0.03	7 (15.6%)	43%	0.72*	0.68
		4. Assessing the costs and resource use of an option	10 (27.8%)	0.26*	0.18	7 (15.6%)	29%	0.73*	0.69
		5. Assessing the efficiency in the use of resources	7 (19.4%)	0.80*	0.76				
		6. Identifying equity, ethical, social, and human rights impact of an option	10 (27.8%)	0.30*	0.22	7 (15.6%)	71%	0.79*	0.76
		7. Assessing the acceptability of an option	10 (27.8%)	0.15	0.06	7 (15.6%)	71%	0.81*	0.78

	C. Maximizing the expected impact of options	1. Adjusting options and enabling factors to maximize impact	8 (24.2%)	0.20	0.09	7 (15.6%)	43%	0.59*	0.53
		2. Finding population groups, settings, and contexts to focusing options	7 (21.2%)	0.11	-0.03	7 (15.6%)	43%	0.62*	0.56
	D. Contributing to prioritize and select options	1. Creating packages of options	7 (25%)	0.51*	0.42	7 (15.6%)	71%	0.76*	0.72
		2. Creating a ranking of options	7 (25%)	0.61*	0.55				
III. Implementing options	A. Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies	1. Identifying and understanding barriers and implementation strategies to deal with them	6 (37.5%)	0.33*	0.19	7 (15.6%)	43%	0.60*	0.53
		2. Identifying and understanding facilitators and implementation strategies to take advantage of them	6 (37.5%)	0.38*	0.26	7 (15.6%)	43%	0.50*	0.41
		3. Prioritizing barriers, facilitators, and implementation strategies	6 (37.5%)	0.40*	0.28	7 (15.6%)	43%	0.51*	0.43
	B. Planning and describing the implementation of an option	1. Identifying who has to do what to implement an option	5 (83.3%)	0.27	0.08	7 (15.6%)	71%	0.84*	0.81
		2. Identifying the context in which the option could be implemented	4 (66.7%)	0.34	0.12	7 (15.6%)	43%	0.75*	0.71
		3. Describing whether implementation of an option is underway and at what stage level	4 (66.7%)	0.30	0.06	7 (15.6%)	43%	0.72*	0.67
IV. Monitoring and evaluation	A. Identifying measurement strategies for populations and outcomes	1. Identifying instruments to ascertain populations	4 (23.5%)	0.24	-0.02	6 (13.3%)	33%	0.67*	0.61
		2. Choosing the most accurate instruments to ascertain populations	5 (29.4%)	0.25	0.07	6 (13.3%)	50%	0.53*	0.44
		3. Identifying measurement instruments for outcomes of interest	4 (23.5%)	0.39	0.18	6 (13.3%)	67%	0.71*	0.65

	4. Determining the best instruments to measure outcomes of interest	5 (29.4%)	0.35	0.19	6 (13.3%)	33%	0.54*	0.45
B. Monitoring and evaluating populations and outcomes of interests	1. Measuring the impact of an option or implementation strategy	6 (27.3%)	0.63*	0.56				
	2. Interpreting the findings of measuring the impact of an option or implementation strategy	5 (22.7%)	0.30	0.12	6 (13.3%)	67%	0.84*	0.808

*: p-value < 0.05
**KW: Kendall's W. Values closer to one means higher consensus rate.
***RC: Spearman rank correlation; Values closer to 1 means higher consensus rate.
NA: Not available (the first round produced too many ties to calculate a valid Kendall's W); Blank cells are questions that were not included in the second round.

S3 table 1. Ranking of study designs per Delphi round in stage I. Clarifying a societal problem, its causes, and potential impacts

Goal	Type of question	Study design	Median rank (IQR) round 1	Median rank (IQR) round 2
A. Choosing and prioritizing outcomes of a problem.	1. Identifying outcomes to characterize a problem	Review of outcomes that have been used by other studies (e.g., scoping review)	1(1,1)	1(1,1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experience	3(3,4)	2(2,4)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of experts' opinion	3(2.5,4)	3(3,3)
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions are currently using	4(2.5,5.5)	4(2.75,4)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	5(3,5.5)	5(5,5)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	5.5(3,5.75)	6(5.875,6)
	2. Understanding individuals' values regarding outcomes	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	2.5(1.25,3)	1(1,1.25)
		Discrete choice experiment (stated preferences)	3.5(1.5,5.125)	2(2,4.625)
		Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	4(3.25,5.125)	4(2.75,4.25)
		Case reports (case series) of people's experience	3.5(2.25,4)	4(3,4.625)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4.5(2.5,5.375)	5(4.375,5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	5.5(2.875,5.5)	5.25(3,6)
	3. Prioritizing outcomes to	Delphi study (to get consensus from experts)	2(1,2)	1(1,1)

	characterize a problem	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	8(3,8)	2(2,2.25)
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions are using	8(3,8)	5(2.75,6)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	8(4,8)	3.5(3,5.25)
		Discrete choice experiment (stated preferences)	8(8,8)	5(4,6.25)
		Ecological study (population-based study, including spatial analysis)	8(4,8)	5.5(5,7.125)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	8(3,8)	7(4.75,8)
		Modelling to compare different measurements	8(4,8)	7(7,7)
		Nominal group technique (NGT)	NA	8.5(7.375,9)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	8(8,8)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	8(8,8)	NA
		Multi-criteria (objective) decision analysis	12(12,12)	NA
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	12(12,12)	NA
B. Describing a problem and its magnitude.	1. Describing a problem in a point in time	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	1(1,1.5)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	3(2.5,3)	NA
		Delphi study (to get consensus from experts)	4(2.5,5)	NA
		Case reports (case series)	6(5,6)	NA
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions are using	6(4,6)	NA
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	6(5,6)	NA

		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	6(4.5,6)	NA
		Modelling to estimate a reality or an indicator	8(8,8)	NA
	2. Describing a problem during a period of time	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	1.5(1,2.25)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	2(1.75,2.25)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	3(2.5,3.25)	NA
		Delphi study (to get consensus from experts)	6.5(5.375,6.625)	NA
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions are using	6.5(5.875,6.625)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	6.5(5.875,6.625)	NA
		Case reports (case series)	6.5(6.125,6.5)	NA
		Modelling to estimate a reality or an indicator	6.5(5.875,6.5)	NA
C. Understanding a problem.	1. Finding conceptual approaches to understand a problem	Review to identify existing frameworks (conceptual analysis)	1(1,3)	1(1,1)
		Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	2(2,3.75)	2(2,3.25)
		Review to build a new framework (critical interpretive synthesis)	2.5(2,4.75)	3(2.75,3)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	3(1.5,3)	4(4,4)
		Descriptive (non-qualitative) case study	5(4.125,5)	5(4.5,5)
	2. Understanding stakeholders' perceptions of a problem	Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	2(2,3)	1(1,1.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	3(1,5.5)	2(2,2)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	4(2,5.5)	3(3,3.25)

		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4(2,5)	4(4,4.25)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4(3,5)	5(4.5,5)
		Discrete choice experiment (stated preferences)	5(3,5.5)	6(5.75,6)
	3. Understanding the role of context in a problem	Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	2(1,3)	1.5(1,3)
		Jurisdictional scan (comparative analysis) to understand the role that the context had in other jurisdictions	6(3,6)	3.5(1,5)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	4(2,6)	2.5(2,5.125)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	5(3,6)	3.5(3,5.625)
		Social network analysis (mapping network analysis)	5(3,6.5)	4(4,5.625)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	6(4,7)	5.5(3.75,6)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	6(2,6.5)	6.75(5,7)
		Descriptive (non-qualitative) case study	6.5(6,6.5)	NA
D. Assessing the variability of a problem.	1. Assessing variability over time	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	2(1.5,3.75)	1(1,1)
		Descriptive (not predicting) time-series analysis (including trend analysis)	3(2,3)	2(2,2)
		Modelling to predict future scenarios (e.g., system dynamics, ARIMA models, etc.)	4(3,4.75)	3(3,4.5)
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	4(3.5,4)	4(3.25,4)
		Case reports (case series)	5.5(3.25,5.5)	5(5,5.75)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	5.5(3.75,5.5)	5.75(4.375,6)
	2. Assessing variability across	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	1.5(1,3.5)	1(1,2.5)

populations and locations	Delphi study (to get consensus from experts on what population is most affected)	4(4,5.25)	2(2,5.375)
	Jurisdictional scan (comparative analysis) to understand variability across populations in other jurisdictions	5(1.75,8.25)	3(3,3)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	5(2,8)	4(4,4)
	Social network analysis (mapping network analysis) to understand people's interactions	5.5(3,8.25)	5(5,6.125)
	Ecological study (population-based study, including spatial analysis)	6.5(4,8)	6(6,6.375)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	8(6.75,8)	7.5(5.5,8)
	Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	7(5.25,8)	7(6.625,7)
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	8(7,8)	9(7.875,9)
	Case-control study (case-comparison study)	12(12,12)	NA
	Case reports (case series)	12(12,12)	NA
	Modelling to estimate an indicator (e.g., the impact of a problem on a given population)	12(12,12)	NA
3. Assessing the importance of a problem relative to other problems	Delphi study (to get consensus from experts)	2.5(1.75,3.75)	1(1,1.75)
	Modelling to estimate an indicator that allows comparisons in common units (e.g., DALYs)	2.5(1,4.75)	2(2,2.75)
	Ecological study (population-based study, including spatial analysis)	4(3.5,5.75)	3(3,4.5)
	Multi-criteria (objective) decision analysis	4.5(2,7)	4(4,5.875)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of quantitative data (e.g., demographic information)	6.5(4.75,7.125)	5(5,5.75)
	Case reports (case series)	6.5(5,7.125)	6.25(6,6.875)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's (affected by the problem) experiences (not asking about hypothetical scenarios)	6.5(5.25,7.125)	7(4,7)

		Discrete choice experiment (stated preferences)	6.5(5.5,7)	7.5(6.625,8)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	6.5(5.75,7)	8.5(7.25,9)
		Descriptive (non-qualitative) case study	10(10,10)	NA
E. Understanding the causes and aggravating factors of a problem.	1. Identifying causes and/or aggravating factors of a problem	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	1(1,2)	2(1,2)
		Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	3(1,4)	3(1,3)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	3(2,8.5)	3(2,4)
		Case-control study (case-comparison study)	4(3,4)	4(4,6.75)
		Interrupted time-series analysis (including joint-point regression)	8.5(8.5,9.5)	6.5(5,7.5)
		Regression discontinuity study (regression kink study or analysis)	8.5(6,8.5)	6.5(6,7.5)
		Ecological study (population-based study, including spatial analysis)	8.5(5,8.5)	7(6.75,7)
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	8.5(8.5,9.5)	7(6.25,9)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	8.5(6,8.5)	7(6.75,8.5)
		Instrumental variables study (two-stage least-squares study or regression)	8.5(8.5,9.5)	8(6.75,8)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	13(13,13)	NA
		Case reports (case series)	13(13,13)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	13(13,13)	NA
	2. Understanding the relative importance of causes and/or	Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	4(3,9.5)	2(1.25,2)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	9(1,9.5)	2.5(2,3.75)

	aggravating factors across population groups	Delphi studies (to get consensus from experts)	9(3,9.5)	3(3,5.25)
		Case-control study (case-comparison study)	3(3,9.5)	5(1.75,6.375)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9(4,9.5)	5.5(5,6)
		Ecological study (population-based study, including spatial analysis)	9(5,9.5)	6(6,6)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	9(2,9.5)	6.75(3.875,7)
		Regression discontinuity study (regression kink study or analysis)	9.5(4,9.5)	6.25(5.25,7.625)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	9.5(9,9.5)	NA
		Multi-criteria (objective) decision analysis	9.5(9,9.5)	NA
		Case reports (case series)	15(15,15)	NA
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	15(15,15)	NA
		Instrumental variables study (two-stage least-squares study or regression)	15(15,15)	NA
		Interrupted time-series analysis (including joint-point regression)	15(15,15)	NA
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	15(15,15)	NA
F. Understanding the impacts of a problem.	1. Identifying impacts/spillover effects of a problem	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3.5(1.5,7.75)	NA
		Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	3.5(3,7.75)	NA
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	5(1,9)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	6.5(2.5,9)	NA
		Interrupted time-series analysis (including joint-point regression)	7(2.75,9)	NA

	Instrumental variables study (two-stage least-squares study or regression)	7(3.5,9)	NA
	Case-control study (case-comparison study)	9(5.25,9.375)	NA
	Modelling to predict future scenarios (e.g., system dynamics, ARIMA models, etc.)	9(6.75,9)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9(9,9.375)	NA
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	9(9,9)	NA
	Ecological study (population-based study, including spatial analysis)	9(9,9)	NA
	Regression discontinuity study (regression kink study or analysis)	9(9,9)	NA
	Case reports (case series)	14(14,14)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	14(14,14)	NA
2. Prioritizing the most important impacts/spillover effects of a problem	Review to find causes or aggravating factors that have been identified by other studies (e.g., scoping review)	4.5(3.25,6.5)	NA
	Delphi studies (to get consensus from experts)	5.5(1,10.375)	NA
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	7.25(2,9.875)	NA
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	7.75(3,9.875)	NA
	Multi-criteria (objective) decision analysis	9.75(5.375, 10.375)	NA
	Instrumental variables study (two-stage least-squares study or regression)	9.75(5.375, 10.375)	NA
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	9.75(4.625, 10)	NA
	Interrupted time-series analysis (including joint-point regression)	9.75(5.375, 10)	NA
	Ecological study (population-based study, including spatial analysis)	10(9.625,10.375)	NA

Modelling to predict future scenarios (e.g., system dynamics, ARIMA models, etc.)	10(9.625,10.375)	NA
Regression discontinuity study (regression kink study or analysis)	10(9.625,10.375)	NA
Case-control study (case-comparison study)	10.25(9.625,10.5)	NA
Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	10.25(10,10.5)	NA
Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	10.25(10,10.5)	NA
Case reports (case series)	16(16,16)	NA
Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	16(16,16)	NA

NA: Not available (questions were not included in the second round).

S3 table 2. Ranking of study designs per Delphi round in stage II. Finding and selecting options to address a problem.

Goal	Type of question	Study design	Median rank (IQR) round 1	Median rank (IQR) round 2
A. Finding and understanding potential options.	1. Scoping a list of potential options	Review to find options that have been used by other studies (e.g., scoping review)	1(1,1)	NA
		Jurisdictional scan (comparative analysis) to understand what options have been implemented by other jurisdictions	2(2,2.25)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions	3.5(2.75,4.25)	NA
		Ecological study (population-based study, including spatial analysis)	3.75(3,4.625)	NA
		Descriptive (non-qualitative) case study	4.25(4,4.625)	NA
	2. Understanding the way potential options and their components work	Randomized-controlled study (randomized experiment or randomized trial) measuring intermediate outcomes	2.5(2,4.875)	1(1,3.5)
		Review to identify existing frameworks (conceptual analysis) that explain how an intervention might work	6.75(1.75,11.5)	2(1,2)
		Interrupted time-series analysis (including joint-point regression) measuring intermediate outcomes	7.25(3.25,10.75)	3(3,4.5)
		Descriptive (non-qualitative) case study	7.25(3.5,10.75)	4(4,4)
		Case reports (case series)	11(8.125,11.5)	5(5,6.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	11(8.625,11.5)	6(6,6.25)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data) measuring intermediate outcomes	11(8.875,11.5)	7(7,7)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	11(8.625,11.5)	8(7,8)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	11.5(8.875,11.5)	9(4.75,9)
		Case-control study (case-comparison study) measuring intermediate outcomes	19(19,19)	NA

		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data) measuring intermediate outcomes	19(19,19)	NA
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs) measuring intermediate outcomes	19(19,19)	NA
		Ecological study (population-based study, including spatial analysis)	19(19,19)	NA
		Instrumental variables study (two-stage least-squares study or regression) measuring intermediate outcomes	19(19,19)	NA
		Modelling to predict the mechanism of action of a given intervention	19(19,19)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	19(19,19)	NA
		Regression discontinuity study (regression kink study or analysis)	19(19,19)	NA
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study) measuring intermediate outcomes	19(19,19)	NA
		Social network analysis (mapping network analysis)	19(19,19)	NA
B. Assessing the expected impact of options.	1. Assessing the feasibility of an option	Delphi studies (to get consensus from experts)	3.5(3,6.25)	1(1,2)
		Jurisdictional scan (comparative analysis) to understand the feasibility of the option elsewhere	5.25(1.25,7.5)	2(1,2)
		Discrete choice experiment (stated preferences)	6.5(3.25,7.5)	3(3,5.75)
		Modelling to predict whether the option will be feasible (e.g., system dynamics, ARIMA models, etc.)	7.25(3.25,7.5)	4(4,6.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	7.25(2.5,7.5)	5(5,6.75)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions (not asking about hypothetical scenarios)	7.25(2,7.5)	6(6,6.25)
		Descriptive (non-qualitative) pilot case study	7.5(3.5,7.5)	7(6.5,7)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	7.5(4.75,7.5)	7(6.5,8)
		Social network analysis (mapping network analysis)	7.5(4.75,7.5)	9(6.75,9)

	Randomized-controlled study (randomized experiment or randomized trial), including pilot RCTs	7.5(7.125,7.5)	NA
2. Assessing the benefits and early-and-frequently occurring harms of an option	Randomized-controlled study (randomized experiment or randomized trial)	1(1,7.25)	1(1,1)
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3(2.25,8.125)	2(2,2)
	Interrupted time-series analysis (including joint-point regression)	7.5(4,9.5)	3(3,3)
3. Assessing late-occurring harms and risks of an option	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	8(4.25,9.5)	4(4,5.5)
	Instrumental variables study (two-stage least-squares study or regression)	8(4.25,9.5)	5(5,5.75)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	8.5(3.75,9.5)	6(5.25,6)
	Regression discontinuity study (regression kink study or analysis)	9.25(3.5,9.5)	7(7,7)
	Modelling to predict or estimate the benefits of an intervention (e.g., system dynamics, ARIMA models, etc.)	9.5(6.125,9.875)	8(8,8)
	Case-control study (case-comparison study)	9.25(9,9.5)	NA
	Ecological study (population-based study, including spatial analysis)	9.5(9.125,9.875)	NA
	Case reports (case series)	9.5(9.5,10)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9.5(9.125,9.875)	NA
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	9.5(9.5,10)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	9.5(9.5,10)	NA
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study), including databases of adverse event reporting (e.g., pharmacovigilance)	4(3,9.375)	1(1,3.5)
	Randomized-controlled study (randomized experiment or randomized trial)	5.5(1,9.5)	2(1,2)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	6.5(2,9.5)	2(2,3)

	Case-control study (case-comparison study)	8(1.75,9.5)	4(4,5.5)
	Instrumental variables study (two-stage least-squares study or regression)	9.25(4,9.5)	5(4,5)
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	9.5(5.25,9.5)	6(6,7)
	Ecological study (population-based study, including spatial analysis)	9.5(4.5,9.5)	7(7,7)
	Interrupted time-series analysis (including joint-point regression)	9.5(6.75,9.5)	8(7,8)
	Case reports (case series)	9.5(8.25,9.5)	9(8,9)
	Modelling to predict or estimate the harms and risks of an intervention (e.g., system dynamics, ARIMA models, etc.)	9.5(9.125,9.5)	NA
	Regression discontinuity study (regression kink study or analysis)	9.5(9.125,9.5)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	9.5(9.125,9.5)	NA
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9.5(9.5,9.5)	NA
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	9.5(9.5,9.5)	NA
4. Assessing the costs and resource use of an option	Modelling to estimate the cost of an option	2(2,3.75)	1(1,1)
	Jurisdictional scan (comparative analysis) to understand the costs in other jurisdictions	9.75(3.25,10.5)	2(2,3)
	Case reports (case series)	6.75(3,10.5)	3(2,7)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9.75(5.25,10.5)	4(4,6)
	Delphi studies (to get consensus from experts)	10(4.5,10.5)	5(5,7.25)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	10.25(5.375,10.5)	6(6,7.25)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	10.25(7.625,10.5)	7(5.5,7.75)

	Instrumental variables study (two-stage least-squares study or regression)	10.25(9.625,10.5)	7.5(7,8.75)
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study), including databases of adverse event reporting (e.g., pharmacovigilance)	10.5(9.625,10.5)	8.5(7.25,9.75)
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	10.5(9.625,10.5)	9(8,9.25)
	Randomized-controlled study (randomized experiment or randomized trial)	10.5(9.625,10.5)	9.5(8,11)
	Case-control study (case-comparison study)	10.5(10.125,10.5)	NA
	Ecological study (population-based study, including spatial analysis)	10.5(10.125,10.5)	NA
	Regression discontinuity study (regression kink study or analysis)	10.5(10.125,10.5)	NA
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	10.5(10.5,10.5)	NA
	Interrupted time-series analysis (including joint-point regression)	10.5(10.5,10.5)	NA
5. Assessing the efficiency in the use of resources	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses)	1(1,1)	NA
	Jurisdictional scan (comparative analysis) to understand whether the option was efficient in other jurisdictions	2(2,2.5)	NA
	Delphi studies (to get consensus from experts)	3(2.5,3)	NA
6. Identifying equity, ethical, social, and human rights impact of an option	Delphi studies (to get consensus from experts)	2(2,3.75)	1(1,1.5)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	2.5(1,3.75)	2(2,2)
	Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	3.5(2,5.875)	3(2.5,3)
	Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4(3,5.875)	4(4,4)
	Descriptive (non-qualitative) case study	5.25(3,5.5)	5(5,5)
	Qualitative inductive (from particular to general i.e., creating theory) methods to interpret/critically analyze a phenomenon (e.g., ethnographic approaches, phenomenology)	5.5(3.25,6)	6(5.75,6)

		Discrete choice experiment (stated preferences)	6(5.625,6)	7(6.25,7)
	7. Assessing the acceptability of an option	Discrete choice experiment (stated preferences)	2.5(1.25,4)	1(1,1.5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	2.5(2,4)	2(2,2)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	3(2,4)	3(3,3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	3(2,3.75)	4(4,4)
		Randomized-controlled study (randomized experiment or randomized trial) measuring people's acceptability	4.75(4,5)	NA
C. Maximizing the expected impact of options	1. Adjusting options and enabling factors to maximize impact	Randomized-controlled study (randomized experiment or randomized trial) to compare different forms of the same intervention	1(1,4.25)	1(1,1)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	4(2,9.5)	2(2,4.25)
		Interrupted time-series analysis (including joint-point regression)	6.75(4,9.5)	4(4,5.5)
		Instrumental variables study (two-stage least-squares study or regression)	5.5(3.75,9.5)	4(3,6.5)
		Randomized-controlled study (randomized experiment or randomized trial) using posthoc comparisons	9.5(2,9.875)	6(3,6)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	9.5(7,9.5)	6(4,8)
		Modelling to predict or estimate the impact of an intervention (e.g., system dynamics, ARIMA models, etc.)	6.75(3,10.25)	6.25(5,7.875)
		Regression discontinuity study (regression kink study or analysis)	9.5(3,9.625)	7(7,7.375)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	9.5(7.5,9.5)	8.5(7.625,9)
		Case-control study (case-comparison study)	9.5(8.375,9.875)	9.5(8.25,10)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9.5(9,9.5)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	9.5(9.125,9.875)	NA

		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	9.5(9.5,9.625)	NA
		Case reports (case series)	9.5(9.5,9.875)	NA
	2. Finding population groups, settings, and contexts to focusing options	Case-control study (case-comparison study)	4(2.5,9.5)	1.5(1,2.75)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	4(2.5,7.25)	2(1.25,2)
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	7(3.5,9.5)	3(3,5.25)
		Randomized-controlled study (randomized experiment or randomized trial) using subgroup comparisons.	9.5(1,9.5)	4(2.5,4)
		Case reports (case series)	9.5(4.5,9.5)	5(5,6.875)
		Instrumental variables study (two-stage least-squares study or regression)	9.5(6.75,9.5)	6(6,7.125)
		Ecological study (population-based study, including spatial analysis)	9.5(5,9.5)	7(7,7.375)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	9.5(2.5,9.75)	7.75(7.125,8)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	9.5(5.75,10.25)	8.5(7.625,9)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	9.5(7,9.5)	8.75(7.125,10)
		Modelling to predict or estimate the impact of an intervention (e.g., system dynamics, ARIMA models, etc.)	9.5(7,9.5)	11(10.25,11)
		Interrupted time-series analysis (including joint-point regression)	9.5(9.5,9.5)	NA
		Regression discontinuity study (regression kink study or analysis)	9.5(9.5,10.25)	NA
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	9.5(9.5,9.5)	NA
D. Contributing to prioritize	1. Creating packages of options	Evidence synthesis of studies evaluating the impact of single interventions to analyze the combined effect of packages.	2(1,3.5)	1(1,1)
		Randomized-controlled study (randomized experiment or randomized trial) to compare packages of interventions in different arms	3(1.5,7.5)	2(2,2)

and select options.	Randomized-controlled study (randomized experiment or randomized trial) using posthoc comparisons.	6(4,11.5)	3(3,3)
	Delphi study (to get consensus from experts)	7(3.5,12)	4(4,4)
	Modelling to predict or estimate the impact of interventions and packages of interventions (e.g., system dynamics, ARIMA models, etc.)	11(4,12)	5(5,5)
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	12(7,12)	6(6,6)
	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses)	12(7,12)	7(7,8.25)
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	12(11.5,12)	8(8,8.75)
	Ecological study (population-based study, including spatial analysis)	12(11.5,12)	9(9,9.25)
	Case-control study (case-comparison study)	12(11.5,12)	10(9.75,10)
	Interrupted time-series analysis (including joint-point regression)	12(10.5,12)	11(10.5,11)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	12(11,12)	12(11,12)
	Discrete choice experiment (stated preferences)	12(11.5,13)	13(11.5,13)
	Multi-criteria (objective) decision analysis	12(11.5,13)	14(12,14)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis)	12(12,12)	NA
	Instrumental variables study (two-stage least-squares study or regression)	12(12,12.5)	NA
	Case reports (case series)	12(12,13)	NA
	Regression discontinuity study (regression kink study or analysis)	12(12,13)	NA
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	12(12,13)	NA
2. Creating a ranking of options	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses) to create a ranked list of options	1(1,2)	NA

Ranking type Delphi study (to get consensus from experts)	2(2,2.5)	NA
Multi-criteria (objective) decision analysis to create a ranked list of options	2(2,3)	NA
Discrete choice experiment (stated preferences)	4(4,4.75)	NA
Jurisdictional scan (comparative analysis) to understand what other jurisdictions have ranked.	4.5(4,5.25)	NA
Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	5.5(5.25,5.75)	NA

NA: Not available (questions were not included in the second round).

S3 table 3. Ranking of study designs per Delphi round in stage III. Implementing or scaling-up an option

Goal	Type of question	Study design	Median rank (IQR) round 1	Median rank (IQR) round 2
A. Setting up a sustainable implementation process by identifying barriers, facilitators, and implementation strategies.	1. Identifying and understanding barriers and implementation strategies to deal with them	Review to find barriers and implementation strategies that have been used by other studies (e.g., scoping review)	1.5(1,2.75)	1(1,1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	3(3,3)	2(2,2.5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	6.5(3,7)	4(3.5,4.75)
		Social network analysis (mapping network analysis) to identify barriers and implementation strategies	4(4,6.25)	5(3,5.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	7(3.25,7)	5(5,5.25)
		Jurisdictional scan (comparative analysis) to understand what barriers and implementation	7(4,7)	6(3,6)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	7(5.5,7)	7(5,7)
		Modelling to predict the mechanism of action of a given barrier or implementation strategy	7(7,7)	NA
		Descriptive (non-qualitative) case study	7(7,7)	NA
	2. Identifying and understanding facilitators and implementation strategies to take advantage of them	Review to find facilitators and implementation strategies that have been used by other studies (e.g., scoping review)	1.5(1,2.75)	1(1,2.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	3(2.25,3.75)	2(2,2.5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	5.5(3.25,7)	3(3,4.25)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	6.5(3,7)	5(3.5,5)
		Social network analysis (mapping network analysis) to identify facilitators and implementation strategies	5.5(4,7)	5(4,5.75)
		Jurisdictional scan (comparative analysis) to understand what facilitators and implementation strategies have been identified by other jurisdictions	7(4,7)	6(3.5,6)

	Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	7(5.5,7)	7(5.75,7)
	Descriptive (non-qualitative) case study	7(7,7)	NA
	Modelling to predict the mechanism of action of a given facilitator or implementation strategy	7(7,7)	NA
3. Prioritizing barriers, facilitators, and implementation strategies	Delphi study (to get consensus from experts)	1.5(1,3.5)	1(1,1.5)
	Discrete choice experiment (stated preferences)	3(2,8.875)	2(2,6.5)
	Multi-criteria (objective) decision analysis	3.5(3,8.875)	3(3,8)
	Cross-sectional study (survey, point-in-time or snapshot study or analysis)	8.25(2.75,11.5)	4(4,7)
	Evidence synthesis of studies evaluating the impact of single interventions to analyze the combined effect of packages.	11(4.875,11.5)	6(3.5,7)
	Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	10.25(4.75, 11.25)	7(5,8)
	Randomized-controlled study (randomized experiment or randomized trial) of implementation issues (e.g., implementation trial)	11(6.375,11.5)	7(7,8)
	Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	11.5(10.75, 11.5)	8(8,8.5)
	Modelling to predict or estimate the importance of barriers, facilitators, and implementation strategies	11.5(10.75, 11.5)	9(8.5,9)
	Economic evaluations (cost-effectiveness, cost-utility, cost-benefit analyses)	11.5(10.75, 11.875)	9(8,13)
	Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	11.5(10.75, 11.5)	9(8.5,12)
	Prospective cohort study of individual-level data (prospective longitudinal or panel study)	11.5(10.75, 11.5)	10(6,10)
	Case-control study (case-comparison study)	11.5(10.75, 11.875)	10(8.5,11)
	Ecological study (population-based study, including spatial analysis)	11.5(11.125,11.5)	NA
	Interrupted time-series analysis (including joint-point regression)	11.5(11.5,11.875)	NA

		Regression discontinuity study (regression kink study or analysis)	11.5(11.5,11.875)	NA
		Instrumental variables study (two-stage least-squares study or regression)	11.5(11.5,11.875)	NA
		Case reports (case series)	11.5(11.5,11.875)	NA
B. Planning and describing the implementation of an option.	1. Identifying who has to do what to implement an option	Delphi studies (to get consensus from experts)	2(2,2)	1(1,1)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	3(1,6)	2(2,2)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	6(5,7)	3(3,4.75)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	6(4,6)	4(4,4)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	6(4,7)	5(5,5.5)
		Social network analysis (mapping network analysis)	6(3,7)	6(6,6)
		Descriptive (non-qualitative) case study	6(4,7)	7(6.75,7)
		Jurisdictional scan (comparative analysis) to understand what other jurisdictions have identified as behavioural variables	7(6,7)	8(8,8)
		Modelling to predict the mechanism of action of a given barrier or implementation strategy	7(6,7)	9(9,9)
	2. Identifying the context in which the option could be implemented	Jurisdictional scan (comparative analysis) to understand what other jurisdictions have identified as contextual variables	2(1,3.5)	1(1,1)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	3(1.75,4)	2(2,3.5)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	4(2,6)	3(3,3.5)
		Delphi studies (to get consensus from experts)	4(2.5,5.25)	4(3,4)
		Qualitative inductive (from particular to general i.e., creating theory) methods to describe a phenomenon (e.g., grounded theory)	4.5(3.75,5.25)	5(5,5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	5(4.25,5.25)	6(6,6)

	Descriptive (non-qualitative) pilot case study	6(5.75,6.25)	7(7,7)
3. Describing whether implementation of an option is underway and at what stage level	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences	2(1,3.625)	1(1,1)
	Descriptive (not predicting) time-series analysis (including trend analysis)	2(2,2)	2(2,3.75)
	Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	3.75(3,4.625)	3(2.5,3)
	Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches, documentary review of public speeches, etc.)	4.25(4,4.625)	4(4,4.75)
	Delphi studies (to get consensus from experts)	4.75(3.625,5)	5(4.75,5)
	Descriptive (non-qualitative) pilot case study	5(3.625,5.625)	6(4.75,6)

NA: Not available (questions were not included in the second round).

S3 table 4. Ranking of study designs per Delphi round in stage IV. Monitoring implementation and evaluating impacts

Goal	Type of question	Study design	Median rank (IQR) round 1	Median rank (IQR) round 2
A. Identifying measurement strategies for populations and outcomes.	1. Identifying instruments to ascertain populations	Review to find measurement strategies that have been used by other studies (e.g., scoping review)	3(1,5)	1(1,3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions on measurement strategies	2(1,3)	2(1,3)
		Jurisdictional scan (comparative analysis) to understand what measurement strategies have been used by other jurisdictions	2(2,2.5)	2(2,2)
		Descriptive (non-qualitative) case study	4(3.75,4)	4(4,5)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4(2.75,5)	5(4,5)
	2. Choosing the most accurate instruments to ascertain populations	Delphi study (to get consensus from experts)	2(2,4)	1(1,1)
		Multi-criteria (objective) decision analysis	3(2,3)	2(2,2)
		Modelling to compare different measurement strategies	5(4,5)	3(3,3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	6.5(3,7.5)	4(4,4)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	6.5(4,8)	5(5,5)
		Case-control study (case-comparison study)	7(6.5,7.5)	6(6,6)
		Randomized-controlled study (randomized experiment or randomized trial)	7.5(6.5,8)	7(7,7)
		Case reports (case series)	7.5(6.5,8)	8(8,8.5)
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	7.5(6.5,8)	9(8.5,9)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	7.5(6.5,8)	10(8.5,10)

	3. Identifying measurement instruments for outcomes of interest	Review to find measurement strategies that have been used by other studies (e.g., scoping review)	1.5(1,2.5)	1(1,1)
		Jurisdictional scan (comparative analysis) to understand what measurement strategies have been used by other jurisdictions	2(1.75,2.25)	2(2,2)
		Descriptive (non-qualitative) case study	3.5(2.75,4)	3(3,3.75)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's opinions on measurement strategies	4(2.5,5)	4(4,4)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	4.5(3.75,5)	5(4.25,5)
	4. Determining the best instruments to measure outcomes of interest	Delphi study (to get consensus from experts)	2(1,4)	1(1,3.25)
		Multi-criteria (objective) decision analysis	3(2,7.5)	2(2,4.25)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	4(2,4)	3(2.25,3)
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	5(2,7.5)	4(3.25,4)
		Modelling to compare different measurement strategies	7.5(3,8)	5.25(5,5.875)
		Case reports (case series)	7.5(5,7.5)	6(5.625,6)
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	7.5(6,7.5)	7(5.875,7)
		Nominal groups technique (NGT)	NA	8(6.5,8)
		Case-control study (case-comparison study)	7.5(7,7.5)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	7.5(7.5,7.5)	NA
		Randomized-controlled study (randomized experiment or randomized trial)	7.5(7.5,8)	NA
B. Monitoring and evaluating	1. Measuring the impact of an option or	Randomized-controlled study (randomized experiment or randomized trial), including pragmatic trials	1(1,1)	NA
		Controlled before-and-after study of aggregated data (including difference-in-differences study and non-equivalent control group designs)	3.5(2.25,4)	NA

populations and outcomes of interests	implementation strategy	Interrupted time-series analysis (including joint-point regression)	3.5(3,4)	NA
		Regression discontinuity study (regression kink study or analysis)	5(2.5,8.625)	NA
		Instrumental variables study (two-stage least-squares study or regression)	7.25(5,9.5)	NA
		Modelling the impact of an intervention on outcomes that are not observable.	7.75(5.25,9.5)	NA
		Case-control study (case-comparison study)	9.5(8.375,9.5)	NA
		Ecological study (population-based study, including spatial analysis)	9.5(9.5,10.25)	NA
		Retrospective cohort study of individual-level data (retrospective or historical longitudinal, or panel study)	9.5(9.5,10.25)	NA
		Case reports (case series)	9.5(9.5,10.25)	NA
		Prospective cohort study of individual-level data (prospective longitudinal or panel study)	9.75(9.5,10.375)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for other purposes (i.e., secondary data)	10(9.5,10.5)	NA
		Single before-and-after study of aggregated data (pre-post or pretest-posttest study)	10(9.5,10.5)	NA
		Cross-sectional study (survey, point-in-time or snapshot study or analysis) of data collected for this purpose (i.e., primary data)	10(9.5,10.5)	NA
	2. Interpreting the findings of measuring the impact of an option or implementation strategy	Cross-sectional study (survey, point-in-time or snapshot study or analysis) of people's experiences (not asking about hypothetical scenarios)	2(1,2)	1(1,1)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe a phenomenon (e.g., qualitative description, narrative approaches)	3(2,3)	2(2,2)
		Descriptive (non-qualitative) case study	3(2,5)	3(3,3)
		Qualitative deductive (from general to particular i.e., testing theory) methods to describe/critically analyze a phenomenon (e.g., qualitative case studies)	4(3,4)	4(4,4)
		Delphi study (to get consensus from experts)	5(3,5.5)	5(5,5)
		Discrete choice experiment (stated preferences)	5(5,5.5)	NA

NA: Not available (questions were not included in the second round).

Acknowledgements

We are grateful to the multiple methodological experts that were part of this study and helped to prioritize the different study designs.

Chapter 4: A living critical interpretive synthesis on the production and dissemination of living evidence syntheses for decision making

4.1. Preface

The COVID-19 pandemic provided an extraordinary case study of research waste. Living evidence syntheses (LESs) have emerged as a critical approach to regularly update a body of evidence addressing a specific question, and investments in an evolving suite of LESs might have helped to avoid the high volume of low quality and almost immediately outdated ‘rapid reviews. This critical interpretive synthesis provides a theoretical framework about the production and dissemination of LESs.

Dr John Lavis and I were responsible for conceiving the idea, and I designed the protocol. Dr. Gordon Guyatt and Dr Arthur Sweetman provided feedback on the protocol. Dr Gordon Guyatt provided critical insights on the thematic categories that were part of the data analysis. In selecting papers and extracting data, I had important support from Qi Wang, Thomas Piggott, Kerry Waddell, and Peter Bragge. Dr. John Lavis, Dr. Gordon Guyatt and Dr. Arthur Sweetman provided feedback on the draft manuscript.

A living critical interpretive synthesis on the production and dissemination of living evidence syntheses for decision-making

Cristián Mansilla*[1,2], camansil@gmail.com
Qi Wang[1,2], wangq87@mcmaster.ca
Thomas Piggott [3,4], tpiggott@peterboroughpublichealth.ca
Peter Bragge[5], peter.bragge@monash.edu
Kerry Waddell [1,2], waddellk@mcmaster.ca
Gordon Guyatt [3] guyatt@mcmaster.ca
Arthur Sweetman[6], arthur.sweetman@mcmaster.ca
John N Lavis[1,2] lavisj@mcmaster.ca

[1] McMaster Health Forum, McMaster University, 1280 Main St W MML-417, Hamilton, Ontario, Canada, L8S 4L6
[2] Health Policy PhD Program, McMaster University, 1280 Main St W 2C Area, Hamilton, Ontario, Canada, L8S 4K1
[3]Department of Health Research Methods Evidence and Impact, McMaster University, 1280 Main St W 2C Area, Hamilton, Ontario, Canada, L8S 4K1
[4] Peterborough Public Health, 185 King Street, Peterborough, Ontario, Canada, K9J 2R8.
[5] Monash Sustainable Development Institute, Monash University, Wellington Rd, Clayton VIC 3800, Melbourne, Australia
[6] Faculty of Social Sciences, McMaster University, 1280 Main St W Kenneth Taylor Hall Rm. 129, Hamilton, Ontario, Canada, L8S 4M4
*Corresponding author

Word count: 3217

4.2. Abstract

Background

The COVID-19 pandemic has had an unprecedented impact in the global population and has also increased research waste. Living evidence syntheses (LESs) seek to regularly update a body of evidence addressing a specific question. During the COVID-19 pandemic, the production and dissemination of LESs emerged as a cornerstone of the evidence infrastructure. This critical interpretive synthesis answers the compass question: What constitutes an LES to support decision-making; when should one be produced, updated, and discontinued; and how should one be disseminated?

Methods

Searches included the Cochrane Library, EMBASE (Ovid), Health Systems Evidence, MEDLINE (Ovid), PubMed, and Web of Science up to 13 July 2022 and included articles that provided any insights on addressing the compass questions on LESs. Articles were selected and appraised, and their insights extracted. An interpretive and iterative coding process was used to identify relevant thematic categories and create a conceptual framework.

Results

Among the 14,022 non-duplicate records identified, 152 publications proved eligible. Most were non-empirical articles, followed by actual LESs; approximately one in three were published in response to the COVID-19 pandemic. The conceptual framework addresses six thematic categories: (1) what is an LES?; (2) what methodological approaches facilitate LESs production?; (3) when to produce an LES?; (4) when to update an LES?; (5) how to make available the findings of an LES?; and (6) when to discontinue updates to an LES?

Conclusion

LESs can play a critical role in reducing research waste and ensuring alignment with advisory and decision-making processes. This critical interpretive synthesis provides relevant insights on how to better organize the global evidence architecture to support their production.

Registration

PROSPERO registration: CRD42021241875

4.3. Contributions to the literature

- The COVID-19 pandemic positioned living evidence syntheses (LESs) as a key feature of the global evidence architecture. This synthesis creates a framework for the production and dissemination of LESs for decision-making.
- Six thematic categories were identified: (1) what is an LES?; (2) what methodological approaches facilitate LESs production?; (3) when to produce an LES, (4) when to update an LES?; (5) how to make available the findings of an LES?; and (6) when to discontinue updates to an LES?.
- This unique conceptual framework can help connect LESs with their role in decision-making processes during health emergencies and in more routine circumstances.

KEYWORDS

Living evidence syntheses; living systematic reviews; evidence-informed health policymaking; decision-making

4.4. Background

The COVID-19 pandemic has had an unprecedented impact the global population. The World Health Organization (WHO) shows that millions of people died since the start of the pandemic n, which is confirmed by the recently estimated excess mortality reported by several countries (Wang et al., 2022). The COVID-19 pandemic is now seen as the global health event with the greatest consequences to the world's health in the last century.

The COVID-19 pandemic not only stressed public-health systems; it also stressed the existing research infrastructure. Before the pandemic, several researchers had shown a significant increase in research outputs, which had escalated to unprecedented levels, with large variability in value (IOANNIDIS, 2016). Research output accelerated further during the COVID-19 pandemic (Kambhampati et al., 2020), creating even bigger challenges with research waste on the one hand and significant gaps from the perspective of decision-makers on the other hand.

In this context, decision-makers have faced difficulties in finding and using the best available research evidence to address the specific challenges they face. Leaving aside the complexity of the issues that the COVID-19 pandemic brought to the fore, decision-makers faced additional complexity in understanding and interpreting the evidence that the COVID-19 pandemic elicited (Vickery et al., 2022).

Living evidence syntheses (LESs) are an approach to regularly updating a body of evidence addressing a specific question. LESs were first described in the literature in 2017 (Elliott, Synnot, Turner, Simmonds, Pearson, et al., 2017), and began being produced by Cochrane (The Cochrane Collaboration, n.d.-a) and

other evidence producers before the COVID-19 pandemic started. During the pandemic, LESs that produced regularly updated summaries of what was known played an important role in informing decisions was known. Thus, the production, dissemination, and use of LESs are now considered a key cornerstone of the global evidence architecture (Pearson, 2021). Given the recency of the prominence of LESs, each of these dimensions requires greater conceptual clarity.

We began this synthesis by using a compass question worded as follow: "What, when and why to produce and disseminate living evidence syntheses for decision-making?" (registered in the PROSPERO record). A compass question can, however, be iteratively adjusted as greater conceptual clarity is gained (Dixon-Woods et al., 2006). The final version of the compass question is as follows: "What constitutes an LES to support decision-making, when should one be produced, updated and discontinued, with what support produced and updated, and how should one be disseminated?"

4.5. Methods

The protocol of this critical interpretive synthesis has been published in PROSPERO (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021241875) and key details follow (Dixon-Woods et al., 2006).

Search methods

To identify potentially relevant documents, the following bibliographic databases were searched:

- Cochrane Library, including CENTRAL (inception to 13 July 2022)
- Health Systems Evidence (inception to 13 July 2022)
- MEDLINE and EMBASE using Ovid (inception to 13 July 2022)
- PubMed (inception to 13 July 2022)
- Web of Science (inception to 13 July 2022).

The electronic database search was supplemented by examining the references of included articles, and evidence syntheses that were captured in the screening process. Additional file 1 describes the search strategies that were used in each database.

Study selection

By identifying or examining relationships among relevant considerations, eligible articles provided insights on the production or dissemination of LESs for decision-making. No restrictions on study design, language, publication type or publication date were applied.

Articles were excluded if they:

- were not LESs, and did not provide insights on LESs;

- were only LESs providing no insights on the production or dissemination of them;
- only provided insights that were restricted to evidence-to-decision aspects of living guidelines.

Duplicates were removed using EndNote® and Covidence®. Two reviewers independently screened titles, abstracts, and full texts, resolving disagreements by a third reviewer; reviewers used Covidence® to conduct this process.

Data extraction

One reviewer extracted the following characteristics from the included articles:

- lead author, month, and year of publication, and citation;
- type of article (LES as declared by the authors; non-LES; empirical article, excluding evidence syntheses; non-empirical article);
- study design and geographical scope for an empirical study, as reported by the authors;
- sector where the article is relevant (following the taxonomy used by the COVID-END inventory of evidence syntheses (Lavis, 2021): clinical management; public-health measures; health-system arrangements; economic and social responses)
- insights addressing the compass question and its components;
- whether or not the article was produced in the context of the COVID-19 pandemic.

Articles could be highlighted as 'highly relevant', meaning that their insights might be too many to extract them in a few sentences.

Studies with more than one publication were managed as follows:

- if a published and a pre-print version was available, the peer-reviewed version of the article was considered for extraction;
- if a full paper was linked to a conference abstract that was captured by a search strategy, the full paper was considered for extraction;
- if a published protocol of an evidence synthesis was available, both the protocol and the published version were considered for extraction;
- if updates of an LESs were available, the latest update was considered for extraction and, if any additional insights were found in older updates, they too were considered for extraction.

A data extraction template was first piloted by two authors, and the full data-extraction process was conducted in Microsoft Excel®.

Quality assessment

Empirical primary studies and evidence syntheses were appraised for their methodological limitations. For primary studies, the mixed methods appraisal tool (MMAT) (Hong et al., 2018) was used, as it allowed appraisal of a broad range of empirical studies. A single reviewer conducted this appraisal.

Evidence syntheses were evaluated using the AMSTAR instrument (Shea et al., 2007). Two reviewers independently conducted this appraisal, discussing any potential conflicts to reach a consensus. When available, the AMSTAR score posted in the COVID-END Inventory, Health Systems Evidence or Social Systems Evidence was used.

Protocols of studies actively underway were not appraised for their methodological limitations.

Data synthesis

Based on the information collected in the data extraction form, each article was classified according to its contributions to addressing the compass question, and whether or not it provided insights about the production and/or dissemination of living evidence syntheses for decision-making.

Based on all the articles considered as eligible, a conceptual framework was created by conducting a narrative synthesis using a coding strategy from the insights coming from the included documents. This coding was conducted in an interpretive and iterative way, starting by the articles classified as highly relevant in the data-extraction stage. Later, insights from articles in each of the draft thematic categories were incorporated in the framework.

To complement the above, a qualitative analysis was conducted based on discussions that were originated in a listserv that is supported by COVID-END, about the role of living reviews in decision-making. These discussions addressed approaches on how to understand LESs, followed by a question on when updates to an LES should be discontinued.

The insights collected from the literature and the list-serv discussion were visually presented in a conceptual framework and were detailed in a set of tables describing the insights collected from the thematic categories that emerged from these data sources.

Living evidence synthesis strategy

This is a living critical interpretive synthesis. The existing criteria for when a living evidence synthesis is needed (Elliott et al., 2017) were met for this critical interpretive synthesis. First, the issue of living evidence syntheses is clearly an ongoing priority for decision-making. Secondly, while the framework

included here is comprehensive, there might be new literature that could lead to adjustments to specific thematic categories, such as new methodological ways to support the production of living evidence syntheses. Finally, at the time of this review, a number of other living evidence syntheses were ongoing and, this might eventually lead to changes in the findings from this critical interpretive synthesis.

The search strategies will be continuously updated every 12 months to check for any potential new articles, and this synthesis will be updated at least three times after its first publication. Insights gained to that point will inform the timing of subsequent updates.

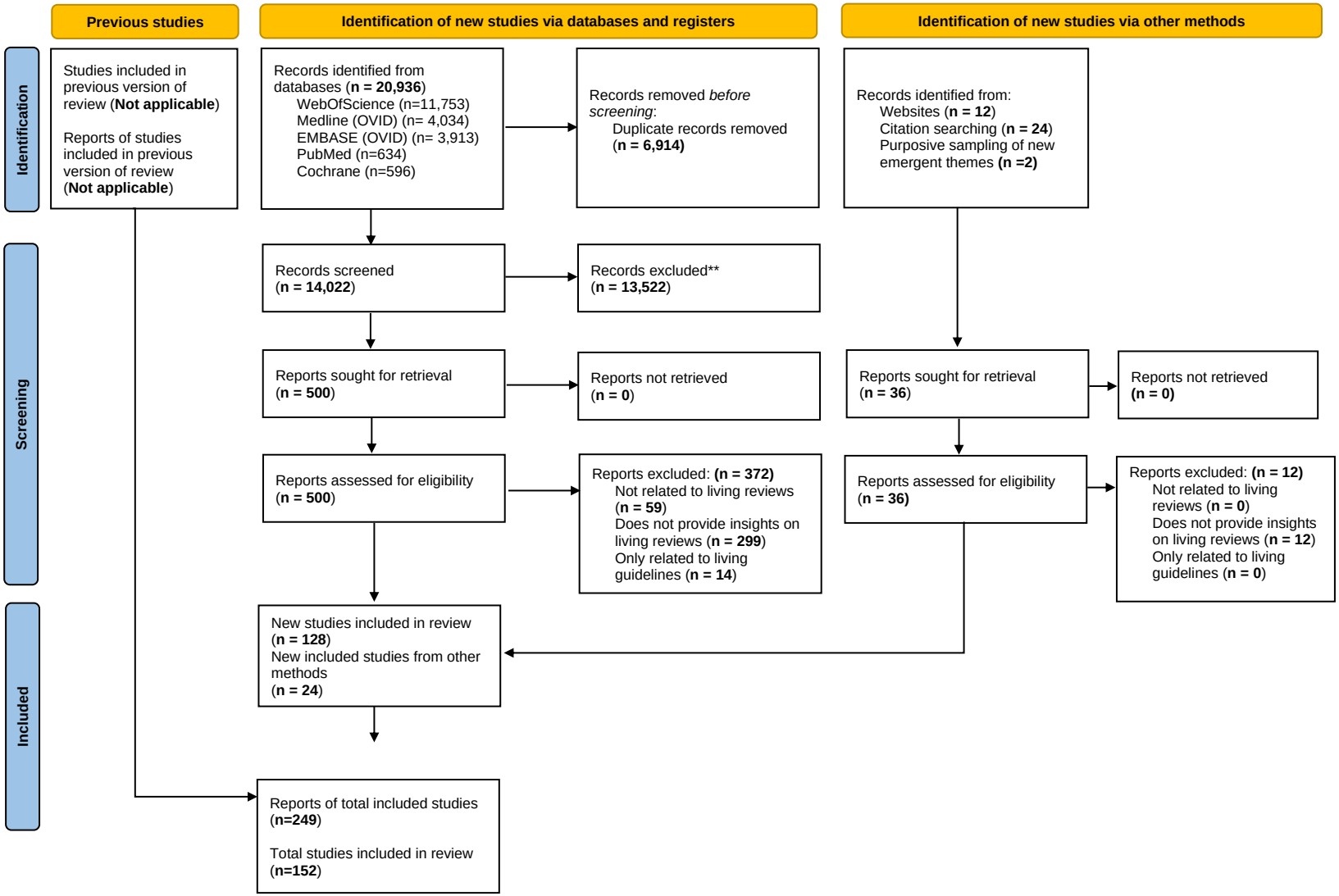
4.6. Results

Search results

Among the 20,936 records found, 14,022 non-duplicated abstracts were screened, and 500 full texts were reviewed for a final set of 128 studies described in 225 publications (many of which were updates of LESs). To fill gaps in the conceptual framework, an additional 24 articles were added using a purposive sampling approach from references of the existing articles, which resulted in 152 included articles. See Figure 1 for the PRISMA flow diagram of the synthesis. Additional File 2 provides a list of the studies excluded in the final stage, along with the reasons for exclusions.

Five conference abstracts proved relevant, but full-text versions of the papers proved unavailable (Britt et al., 2021; Hearnden et al., 2019; Riaz, Siddiqi, et al., 2019; Richard et al., 2020; Siqueira et al., 2021); they were included and extracted in their abstract form.

Figure 4.1. PRISMA diagram showing the review process for selecting the included studies.



Description of studies

Table 1 describes the characteristics of the included studies. The majority of articles that provided insights for this synthesis were evidence syntheses (living or not), followed by non-empirical articles. Only a small number of articles were empirical studies that were not evidence syntheses, and 36% of all studies included were produced in response to the COVID-19 pandemic.

In terms of thematic focus, half of the articles addressed clinical management issues, followed by not having a particular focus and public-health measures. Five articles addressed health-system arrangements and economic and social responses. Finally, thematic categories were relatively equally served in terms of the number of articles, with the exception being when to discontinue updates to an existing LES, an issue addressed by 20 articles (13.2%).

Table.1 Description of the included studies

		N	%
Type of article			
	Living evidence synthesis	57	38
	Non-living evidence synthesis	20	13
	Empirical article (not evidence synthesis)	11	7
	Non-empirical article	62	41
Produced in response to the COVID-19 pandemic		54	36
Sector that is most relevant*			
	Public-health measures	20	13
	Clinical management	79	52
	Health-system arrangements	1	0.7
	Economic and social responses	4	3
	No particular focus	59	39
Thematic categories*			
	Definition of an LES	69	46
	Methods to assess the need and produce an LES	81	53
	When to produce an LES	60	40
	When to update an LES	63	42
	Dissemination of LES findings	49	33
	When an LES is no longer needed	20	13

*One article could address more than one thematic category or sector. Percentages could sum more than 100%

Among the small number of empirical studies, the study designs varied (mixed-method, qualitative, and quantitative); one study was conducted in each of Australia (Turner et al., 2022), Italy (Arienti et al., 2022), and the United States (Kingdon, 2011), and the fourth was conducted in both Australia and Canada (Lee et al., 2022), while the remainder not having a specific geographical scope. Of the 151

articles, 64 were classified as highly relevant mainly based on the importance of their contributions to creating the conceptual framework.

The quality of the evidence syntheses was moderate to high; most proved of as moderate quality in the AMSTAR instrument (4 to 7). The limited number of empirical primary studies showed a wide variation in terms of their methodological limitations but, with the exception of one article, they fill most of the criteria from MMAT. Additional File 3 shows the detail of the AMSTAR scores for evidence syntheses, and methodological limitations for empirical primary studies using the MMAT.

Results of the coding

Six thematic categories were identified from the data sources within which there proved 21 different sub-themes. With the exception of the sub-theme 'labelling living' in the thematic category 1 that emerged only from the listserv discussion, remainder of the thematic categories and sub-theme emerged from both the literature and the listserv discussion.

In conducting the critical analysis, two specific topics emerged as potential controversies or gaps in the literature. First, the definition of what should be considered an update (thematic category 1, sub-theme 2. Updates) was drop by multiple ideas. Secondly, three specific gaps were found in the critical analysis, and they were filled by purposively sampled literature: (1) when an update was needed, which was filled by literature about when a non-living evidence synthesis needs to be updated; (2) when an issue was a priority for decision-making, which was filled by the agenda setting literature; and (3) the applicability of the findings of a living evidence synthesis for different contexts and issues. Complementary, table 2 explains how each sub-theme relates to each thematic category, as well as the citation of the papers contributing to each thematic category. Additional File 4 provides a detailed

description of how each article contributed to each thematic category and sub-theme, while Additional File 5 provides a more thorough description of each one of the thematic categories.

Table 4.2. Description of the thematic categories and sub-themes that emerged from the literature.

	Thematic categories					
	1. Definition of an LES	**2. Methods to assess the need and produce an LES**	**3. When to produce an LES**	**4. When to update an LES**	**5. Dissemination of the findings of an LES**	**6. When to discontinue updates to an LES**
Sub-themes	1.1. LESs: Understanding a living synthesis as a summary of all existing research that is up-to-date at any defined point in time 1.2. Updates: Understanding what constitutes an 'update' in the context of an LES, and the potential changes that it might produce in the existing body of evidence 1.3. Labeling 'living': What do we understand by the label 'living' (approaches to how and when is appropriate to label an evidence synthesis as 'living')	2.1. Assessment of the need of an LES: Methods to predict whether new literature might change the findings (assessing the probability that new evidence changes the findings), the susceptibility of the existing findings, or the value that new information would provide in reducing uncertainty, or how the context and issue might change the applicability of the findings) 2.2. Team management: Methods to facilitate team management (facilitating work in teams and ensuring team sizes) while producing an LES	3.1. Types of decision: Alternatives for an evidence producer when starting a new evidence synthesis (starting a new synthesis, updating one, or making one living) 3.2. Demand-side triggers: 'triggers' when an LES might be needed related to the relevancy of the topic for decision-making (i.e.., priority issue) 3.3. Supply-side triggers: 'triggers' when an LES might be needed related to the susceptibility of the existing evidence to be changed and the probability that new evidence might change the findings. 3.4. Other elements to	4.1. Processed involved: Parts of an evidence synthesis that could be updated, including search methods, data synthesis and publication. 4.2. Demand-side triggers: 'triggers when an LES might need to be updated, related to a change in the priority of an issue for decision-making 4.3. Supply-side triggers: 'triggers when an LES might need to be updated related to a change in the susceptibility of the existing evidence to be changed and the probability that new emerging evidence might change the findings	5.1. Platforms: different options of platforms that could be used to make available the findings (website, scientific journal, interactive platforms), and how they can respond to the multiple challenges of LESs 5.2. Structured format: different adaptations to the format of a LESs that can be used to streamline dissemination processes, including a 'changes from last version' section, highlighting changes in the text, and detailing update plans 5.3. LES users: types of decision-makers and evidence intermediaries that can use the findings of LESs, and	6.1. Demand-side triggers: 'triggers' when updates to an LES could be discontinued because the issue is no longer a priority, or the research question could be re-framed 6.2. Supply-side triggers: 'triggers' when updates to an LES could be discontinued because no new evidence is expected to be available or the findings of the existing evidence are unlikely to change 6.3.Other elements to consider: other elements to be considered when making a decision of when a LES can stop being updated,

		2.3. Production: Methods to facilitate the production (searching, selecting studies, extracting data, assessing risk-of-bias and synthesizing data) of an LES	consider other elements to be considered when making a decision of when an LES needs to be conducted (e.g., workload, context, etc.)	4.4. Frequency: deciding how frequent an LES needs to be updated, with what support researchers count to make this decision, and the need for researchers to commit to reliable schedules for updates	the need for evidence producer to tailor the presentation depending on the type of user. 5.4. Speeding-up: strategies that can be used to reduce the time from which findings are available and they are used by decision-makers and evidence intermediaries, including pre-prints, small-indexed publications notifying updates, among others.	including the engagement of the synthesis team, and the planned obsolescence of an LES.
Citations	(Akl et al., 2020; Bell & Wade, 2021; Boutron et al., 2020; Crequit et al., 2016, 2019, 2020; Donoghue et al., 2018; Dzinamarira et al., 2021; Elliott et al., 2014; Elliott, Synnot, Turner, Simmonds, Akl, et al., 2017; Eshun-Wilson et al., 2019; Franco & Sguassero, 2020; Garner et al., 2016; Hazlewood et al., 2020; Ipekci et	(Ahmadzai et al., 2013; Barrowman et al., 2003; Boutron et al., 2020; Cohen et al., 2009, 2012; Crequit et al., 2016, 2020; Dalal et al., 2013; Donoghue et al., 2018; Dzinamarira et al., 2021; Elliott et al., 2014; Elliott, Synnot, Turner, Simmonds, Akl, et al., 2017; France et al., 2016; Franco & Sguassero, 2020;	(Akl et al., 2020; Boutron et al., 2020; Crequit et al., 2020; Dzinamarira et al., 2021; Elliott et al., 2014; Elliott, Synnot, Turner, Simmonds, Akl, et al., 2017; Eshun-Wilson et al., 2019; Garner et al., 2016; Hazlewood et al., 2020; Ipekci et al., 2021; John et al., 2020; Kingdon, 2011; LIvE Framework, 2020; Macdonald et al., 2020;	(Ahmadzai et al., 2013; Barrowman et al., 2003; Cohen et al., 2009, 2012; Crequit et al., 2019, 2020; Dalal et al., 2013; Eshun-Wilson et al., 2019; France et al., 2016; French et al., 2005; Garner et al., 2016; Ipekci et al., 2021; John et al., 2020; Kingdon, 2011; Korang et al., 2020; Macdonald et al., 2020; Moher et al., 2007, 2008;	(Boutron et al., 2020; Crequit et al., 2016, 2019; Elliott et al., 2014; Elliott, Synnot, Turner, Simmonds, Akl, et al., 2017; Franco & Sguassero, 2020; Garner et al., 2016; Khamis et al., 2019; Korang et al., 2020; LIvE Framework, 2020; Macdonald et al., 2020; Martínez García et al., 2017; Mavergames & Elliott, n.d.; Riaz et	(Boutron et al., 2020; Crequit et al., 2020; France et al., 2016; Hazlewood et al., 2020; Kingdon, 2011; Macdonald et al., 2020; Mavergames & Elliott, n.d.; The Cochrane Collaboration, n.d.)

	al., 2021; John et al., 2020; Juul et al., 2020; Khamis et al., 2019; Korang et al., 2020; Lansky & Wethington, 2020; Lerner et al., 2019; LIvE Framework, 2020; Macdonald et al., 2020; Maguire & Guerin, 2020; Martínez García et al., 2017; Mavergames & Elliott, n.d.-a; Moher & Tsertsvadze, 2006; Mondello et al., 2021; Nikolakopoulou, Egger, et al., 2018; Rahal et al., 2016; Riaz et al., 2021; Riaz, Rawal, et al., 2019; Ritch, 2016; Santillan-Garcia, 2020; Santillan-Garcia et al., 2020; Shokraneh & Russell-Rose, 2020; Siemieniuk et al., 2020; G. K. P. Spurling et al., 2017; The Cochrane	Garner et al., 2016; Juul et al., 2020; Lavis et al., 2009; LIvE Framework, 2020; Maguire & Guerin, 2020; Martínez García et al., 2017; Mavergames & Elliott, n.d.; Moher et al., 2007; Nikolakopoulou, Egger, et al., 2018; Nikolakopoulou, Mavridis, et al., 2018; Riaz et al., 2021; Shekelle et al., 2017; Sutton et al., 2009; Takwoingi et al., 2013; The Cochrane Collaboration, n.d.; Vandvik et al., 2016; Vergara-Merino et al., 2020; Vogel et al., 2019; Wallace et al., 2012; Winters et al., 2021)	Nikolakopoulou, Egger, et al., 2018; Siemieniuk et al., 2020; G. K. Spurling et al., 2017; Vergara-Merino et al., 2020; Vogel et al., 2019; Winters et al., 2021)	Nikolakopoulou, Mavridis, et al., 2018; Shokraneh & Russell-Rose, 2020; G. K. Spurling et al., 2017; Sutton et al., 2009; The Cochrane Collaboration, n.d.; Vergara-Merino et al., 2020; Vogel et al., 2019; Winters et al., 2021)	al., 2021; Ritch, 2015; Shanahan, 2015; Shokraneh & Russell-Rose, 2020; Siemieniuk et al., 2020; Takwoingi et al., 2013; The Cochrane Collaboration, n.d.; Vandvik et al., 2016; Vergara-Merino et al., 2020; Vogel et al., 2019; Winters et al., 2021)	

	Collaboration, n.d.-b; Vandvik et al., 2016; Vergara-Merino et al., 2020; Vogel et al., 2019; Winters et al., 2021)					

LES: Living evidence synthesis

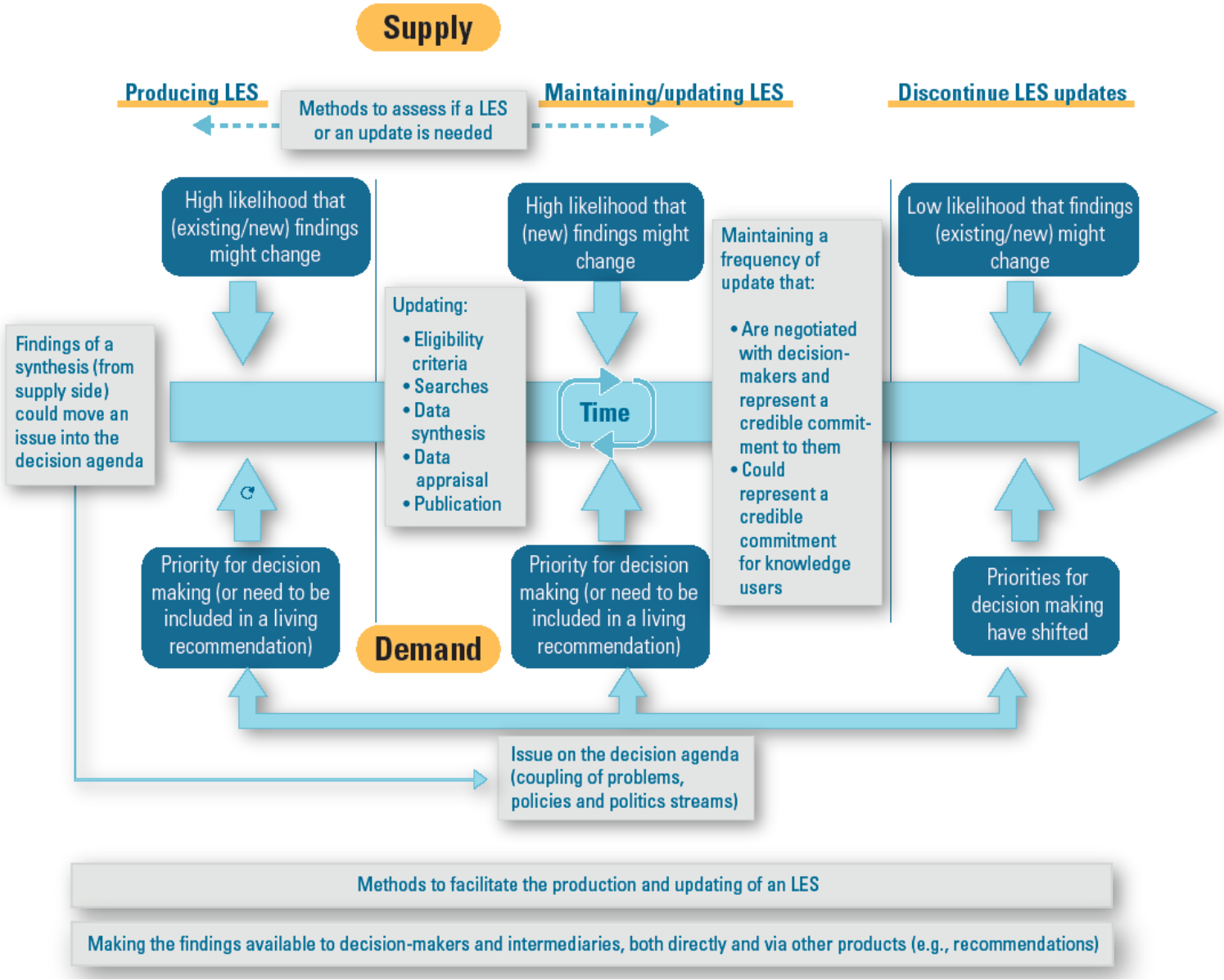
The six thematic categories include 21 sub-themes. In the first thematic category, the definition of LES is separated into what constitutes a living synthesis, an update on the meaning of the label "living". The second thematic category explains the methods that can be used to assess the need for an LES, how to manage a team conducting an LES, and the methods to facilitate the production of an LES. The third and fourth thematic categories include 'triggers' to look for when deciding to produce and update an LES, which are structured into demand-side, supply-side, and other type of triggers. The fifth thematic category describes the platforms and format that an LES can use to disseminate its findings. It also describes the potential users to whom the findings of an LES would be disseminated, as well as ways to speed-up the dissemination of LESs. Finally, the sixth thematic category include 'triggers' to look for when deciding to discontinue updates of an LES.

Conceptual framework

Figure 2 shows the conceptual framework created from the thematic categories found in this critical interpretive synthesis. It displays the three main sections of the cycle of an LES (producing, maintaining/updating, and discontinuing updates), which are described in thematic categories 3, 4 and 6. These three sections are arranged around a time axis from left to right, while this axis divides the supply triggers coming from the upper part of the diagram from the demand triggers coming from its lower part.

The figure also reflects insights gathered from thematic categories 2 (methods) and 5 (making findings available) as cross-cutting topics across the figure. The demand side is mainly driven by how issues are sitting on the decision agenda, using the Kingdon model of agenda setting, by coupling the streams of problems (i.e., why the problem come to attention), policies (a potential viable solution) and politics (political climate that could be conducive) (Kingdon, 2011).

Figure 4.2. Conceptual framework showing demand and supply triggers in three main stages of LES.



This conceptual framework for producing and making available the findings of LES acknowledges that findings of a given synthesis could also contribute to making issues coming to attention, creating a type of feedback. Also, it shows that the conception of an update could come not only from adding new evidence, but also any changes in the underlying structure of an existing synthesis (e.g., eligibility criteria, presentation details, etc.). Additionally, the frequency of updates could be tailored or established in advance, but a negotiation with potential decision-makers and evidence intermediaries is also flagged as important insight, because as long as updating frequently is critical, creating a credible commitment with knowledge users in terms of when to expect new updates is also important. Finally, one important insight gathered from the literature is the fact that the decision of when to start an LES could be similar to the decision regarding when to update one, since every LES will start by a 'baseline' synthesis that will be updated regularly. The framework shows a cycle in terms of the need of assessing when to update an LES.

4.7. Discussion

Principal findings

This critical interpretive synthesis considered a broad literature and a series of posts included in a listserv discussion to create a conceptual framework to understand what LESs are, and when and how to produce and disseminate them. The resulting framework (Figure 2) structured the LES process in three main 'buckets'; starting an LES, its maintenance or updating, and the decision to discontinue updates. It also highlights the main triggers that could inform each stage from the demand and supply sides. While the triggers from the demand side are mainly associated with whether an issue is a priority for decision-making, the triggers from the supply side are associated with the likelihood that the existing body of evidence for a given question might change.

The six thematic categories included 21 sub-themes that were included as part of the analysis reflecting the complexity and the number of different aspects involved in the production and dissemination of LESs. Considering that the first paper on LESs was only published in 2017 (Elliott, Synnot, Turner, Simmonds, Pearson, et al., 2017), this area has grown substantially in complexity. It has also been powered by the COVID-19 pandemic, which established LESs as a key cornerstone of the global evidence architecture (Pearson, 2021).

Findings in relation to other studies

This is the first paper creating a conceptual framework to support the production and dissemination of LESs. While the first paper in this area was published a number of years ago (Elliott, Synnot, Turner, Simmonds, Akl, et al., 2017e), several efforts have been conducted since

then by several evidence producers, including the Cochrane Collaboration (Cochrane Collaboration, 2017). During the COVID-19 pandemic, the number of LESs grew exponentially (Pearson, 2021b), although most were efforts that could be relied on in terms of regular updates (and some of them never made it beyond the publication of a protocol to the publication of their first version). The COVID-19 pandemic highlighted the importance of LESs for decision-making as topics, issues and priorities constantly evolved, as did the evidence production as well.

Strengths and limitations

This paper has important strengths. First, although the main body of literature came from the health sector, it provides a conceptual framework that different types of decision-makers in different sectors. Secondly, it is designed to be a living CIS that will be updated as soon as new literature provides new insights, keeping the conceptual framework always up-to-date. Thirdly, the data sources included were exhaustive, using a comprehensive search of the literature combined with an analysis of dedicated insights on the role of LESs for decision-making. Finally, it incorporates other conceptual frameworks where relevant (e.g., agenda setting processes), providing a more comprehensive understanding of the complex processes addressed.

This article has limitations. First, this paper focused on the production and dissemination of the findings of an LES. Although the potential uses of LESs for decision-making were partially addressed by considering the demand-side triggers to gather emerging insights from the literature, these were beyond the scope of this paper, but there may be insights from that area that may provide insights on this topic. Secondly, we mainly found literature that is not empirical. No rigorous evaluations were available that could address the impact of LESs for decision-making.

Finally, some parts of the evidence synthesis process were conducted by using only a single reviewer (i.e., data extraction and assessing the methodological limitations of the included articles).

Implications for policy and practice

This framework can inform decision-makers, evidence intermediaries and evidence producers regarding the role that LESs can play in decision-making processes. On the one hand, LESs can inform decision-makers regarding the importance on using LESs for their decisions, as well as considerations related to commissioning and setting expectations for LES teams. On the other hand, it can help evidence intermediaries and producers with demand-side considerations related to conducting, updating, or discontinuing updates to an LES, as well as what methods they can use to facilitate this work.

Evidence producers can use this framework to inform their efforts regarding when to produce an LES. This could help to reduce research waste by facilitating coordination among evidence producers to encourage the production of a suite of high-quality living evidence syntheses on key priority topics, as opposed to have multiple (sometimes duplicate) initiatives conducting non-living evidence syntheses. However, incentives from funders and academic publications might act as a barrier to reach this goal.

When conducting a living evidence synthesis, evidence producers should transparently report their plans regarding update frequencies and how they are planning to be updated. This will help to focus their research funding efforts on topics that would produce sound and relevant LESs.

Additionally, this framework can also be used to consider whether living datasets could also be served by this analysis. Hence, the role of living evidence might not necessarily be at the level of syntheses or documents, but also at other forms of evidence.

Implications for future research

Future research efforts should address how LESs could be better structured and organized by researchers, intermediaries, and decision-makers to better coordinate their actions to facilitate the effective uses of the different types of evidence in the decision-making process. Empirical studies that can ask decision-makers, evidence intermediaries and researchers about how to advance the usefulness of this framework could provide additional insights by conducting prioritizing exercises (e.g., Delphi studies) or providing qualitative insights (e.g., case study) to test to support evidence producers and intermediaries on when to produce, update and discontinue LESs.

4.8. Conclusion

This critical interpretive synthesis provided a thorough conceptual framework to better understand what LESs are, and when and how to produce and disseminate an LES. Six thematic categories emerged from the literature, highlighting definitions and methods to produce an LES, triggers from the demand and supply side to initiate production, update and discontinue updating LESs, and insights to how to make available findings of an LES. This framework can inform decision-makers, evidence intermediaries and evidence producers to clarify the role that LESs can play in decision-making processes. Future research could advance the usefulness of this framework by testing it and putting it in practice to facilitate the use of LESs in decision-making processes.

LIST OF ABBREVIATIONS

CIS: Critical interpretive synthesis

LES: Living evidence synthesis

4.9. Declarations

Ethics approval and consent to participate.

Not applicable

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

The author(s) received no specific funding for this work.

Authors contributions

CM and JNL conceived the idea. CM, AS, GG and JNL wrote the protocol. CM, QW, TP, KW and PB participated in the screening and data extraction. CM performed data analysis and led the article writing, with inputs from JNL, AS and GG. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge and thank all the individuals that participated in the COVID-END listserv discussion on living evidence syntheses.

4.10. References

Ahmadzai, N., Newberry, S. J., Maglione, M. A., Tsertsvadze, A., Ansari, M. T., Hempel, S., Motala, A., Tsouros, S., Schneider Chafen, J. J., Shanman, R., Moher, D., & Shekelle, P. G. (2013). A surveillance system to assess the need for updating systematic reviews. *Systematic Reviews*, *2*(1), 104. https://doi.org/10.1186/2046-4053-2-104

Akl, E. A., Haddaway, N. R., Rada, G., & Lotfi, T. (2020). Future of Evidence Ecosystem Series: Evidence synthesis 2.0: When systematic, scoping, rapid, living, and overviews of reviews come together. *Journal of Clinical Epidemiology*, *123*(PG-162-165), 162–165. https://dx.doi.org/10.1016/j.jclinepi.2020.01.025

Arienti, C., Campagnini, S., Brambilla, L., Fanciullacci, C., Lazzarini, S. G., Mannini, A., Patrini, M., & Carrozza, M. C. (2022). The methodology of a "living" COVID-19 registry development in a clinical context. *Journal of Clinical Epidemiology*, *142*, 209–217. https://doi.org/10.1016/j.jclinepi.2021.11.022

Barrowman, N. J., Fang, M., Sampson, M., & Moher, D. (2003). Identifying null meta-analyses that are ripe for updating. *BMC Medical Research Methodology*, *3*(1), 13. https://doi.org/10.1186/1471-2288-3-13

Bell, V., & Wade, D. (2021). Mental health of clinical staff working in high-risk epidemic and pandemic health emergencies a rapid review of the evidence and living meta-analysis. *Social Psychiatry and Psychiatric Epidemiology*, *56*(1 PG-1–11), 1–11. https://dx.doi.org/10.1007/s00127-020-01990-x

Boutron, I., Ravaud, P., Crequit, P., Williams, H. C., Meerpohl, J., & Craig, J. C. (2020). Future of evidence ecosystem series: 3. From an evidence synthesis ecosystem to an evidence

ecosystem. *Journal of Clinical Epidemiology*, *123*(PG-153-161), 153–161. http://dx.doi.org/10.1016/j.jclinepi.2020.01.027

Britt, A., Yang, E., Crittenden, D., Bhangdia, T., Nye, B., Duffy, E.-C., Miriyala, S., van der Veen, E., Marchant, S., Fan, K., Strauss, E., McKinley, K., Sharda, K., Tuor, M., Mahajan, I., Ranganathan, N., Perriello, L., Wallace, B., Mangalath, P., … Del Vecchio Fitz, C. (2021). Abstract P04: Effects of COVID-19 treatments on cancer: A machine learning approach to synthesize clinical evidence at scale. *Clinical Cancer Research*, *27*(6_Supplement), P04–P04. https://doi.org/10.1158/1557-3265.COVID-19-21-P04

Cochrane Collaboration. (2017). *Living Evidence Network Governance Structure*. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/Living%20Evidence%20Network%20-%20Governance%20Structure_1.pdf

Cohen, A. M., Ambert, K., & McDonagh, M. (2009). Cross-Topic Learning for Work Prioritization in Systematic Review Creation and Update. *Journal of the American Medical Informatics Association*, *16*(5), 690–704. https://doi.org/10.1197/jamia.M3162

Cohen, A. M., Ambert, K., & McDonagh, M. (2012). Studying the potential impact of automated document classification on scheduling a systematic review update. *BMC Medical Informatics and Decision Making*, *12*(1), 33. https://doi.org/10.1186/1472-6947-12-33

Crequit, P., Boutron, I., Meerpohl, J., Williams, H. C., Craig, J., & Ravaud, P. (2020). Future of evidence ecosystem series: 2. Current opportunities and need for better tools and methods. *Journal of Clinical Epidemiology*, *123*(PG-143-152), 143–152. https://dx.doi.org/10.1016/j.jclinepi.2020.01.023

Crequit, P., Martin-Montoya, T., Attiche, N., Trinquart, L., Vivot, A., & Ravaud, P. (2019). Living network meta-analysis was feasible when considering the pace of evidence generation. *Journal of Clinical Epidemiology*, *108*(PG-10-16), 10–16. https://dx.doi.org/10.1016/j.jclinepi.2018.12.008

Crequit, P., Trinquart, L., Yavchitz, A., & Ravaud, P. (2016). Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: The example of lung cancer. *BMC Medicine*, *14*(PG-8), 8. https://dx.doi.org/10.1186/s12916-016-0555-0

Dalal, S. R., Shekelle, P. G., Hempel, S., Newberry, S. J., Motala, A., & Shetty, K. D. (2013). A Pilot Study Using Machine Learning and Domain Knowledge to Facilitate Comparative Effectiveness Review Updating. *Medical Decision Making*, *33*(3), 343–355. https://doi.org/10.1177/0272989X12457243

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, *6*(1), 35. https://doi.org/10.1186/1471-2288-6-35

Donoghue, E., Lunny, C., Synnot, A., Bragge, P., Menon, D., Clavisi, O., Pattuwage, L., Volovici, V., Mondello, S., Cnossen, M. C., Gruen, R. L., & Maas, A. (2018). The currency, completeness, and quality of systematic reviews of acute management of moderate to severe traumatic brain injury: A comprehensive evidence map. *PloS One*, *13*(6 PG-e0198676), e0198676. http://dx.doi.org/10.1371/journal.pone.0198676

Dzinamarira, T., Mhango, M., Dzobo, M., Ngara, B., Chitungo, I., Makanda, P., Atwine, J., Nkambule, S. J., & Musuka, G. (2021). Risk factors for COVID-19 among healthcare

workers. A protocol for a systematic review and meta-analysis. *PloS One*, *16*(5 PG-e0250958), e0250958. https://dx.doi.org/10.1371/journal.pone.0250958

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., & Thomas, J. (2017). Living systematic review: 1. Introduction-the why, what, when, and how. *Journal of Clinical Epidemiology*, *91*(PG-23-30), 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., … Pearson, L. (2017). Living systematic review: 1. Introduction—The why, what, when, and how. *Journal of Clinical Epidemiology*, *91*, 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., & Gruen, R. L. (2014). Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine*, *11*(2 PG-e1001603), e1001603. https://dx.doi.org/10.1371/journal.pmed.1001603

Eshun-Wilson, I., Jaffer, S., Smith, R., Johnson, S., Hine, P., Mateo, A., Stephani, A.-M., & Garner, P. (2019). Maintaining relevance in HIV systematic reviews: An evaluation of Cochrane reviews. *Systematic Reviews*, *8*(1 PG-46), 46. https://dx.doi.org/10.1186/s13643-019-0960-5

France, E. F., Wells, M., Lang, H., & Williams, B. (2016). Why, when and how to update a meta-ethnography qualitative synthesis. *Systematic Reviews*, *5*(1), 44. https://doi.org/10.1186/s13643-016-0218-4

Franco, J. V. A., & Sguassero, Y. (2020). Pandemia por COVID-19 e infodemia: Retos y oportunidades para la síntesis confiable y actualizada del conocimiento reliable and updated synthesis of knowledge. *Revista Argentina de Salud Pública*, *12*(suppl 1 PG-8-8), 8.

French, S. D., McDonald, S., McKenzie, J. E., & Green, S. E. (2005). Investing in updating: How do conclusions change when Cochrane systematic reviews are updated? *BMC Medical Research Methodology*, *5*(1), 33. https://doi.org/10.1186/1471-2288-5-33

Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Schünemann, H. J., Akl, E. A., Beyene, J., Chang, S., Churchill, R., Dearness, K., Guyatt, G., Lefebvre, C., Liles, B., Marshall, R., Martínez García, L., Mavergames, C., Nasser, M., Qaseem, A., Sampson, M., … Wilson, E. C. (2016). When and how to update systematic reviews: Consensus and checklist. *BMJ*, i3507. https://doi.org/10.1136/bmj.i3507

Hazlewood, G. S., Whittle, S. L., Kamso, M. M., Akl, E. A., Wells, G. A., Tugwell, P., Thomas, M., Lee, C., Ejaredar, M., Choudhary, D., & et al. (2020). Disease-modifying anti-rheumatic drugs for rheumatoid arthritis: A systematic review and network meta-analysis. *Cochrane Database of Systematic Reviews*, *3 PG-*. https://doi.org/10.1002/14651858.CD013562

Hearnden, J., Dudoit, K., Kim, E., Tremblay, G., & Forsythe, A. (2019). PMU118 USE OF COMPUTER-ASSISTED METHODS TO REALIZE THE CONCEPT OF A LIVING SYSTEMATIC REVIEW VIA AN ONLINE PLATFORM. *Value in Health*, *22*, S729. https://doi.org/10.1016/j.jval.2019.09.1736

Hong, Q. N., Pluye, P., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-C., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I.

(2018). *Mixed methods appraisal tool (MMAT) version 2018*. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf

IOANNIDIS, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Ipekci, A. M., Buitrago-Garcia, D., Meili, K. W., Krauer, F., Prajapati, N., Thapa, S., Wildisen, L., Araujo-Chaveron, L., Baumann, L., Shah, S., Whiteley, T., Solis-Garcia, G., Tsotra, F., Zhelyazkov, I., Imeri, H., Low, N., & Counotte, M. J. (2021). Outbreaks of publications about emerging infectious diseases: The case of SARS-CoV-2 and Zika virus. *BMC Medical Research Methodology*, *21*(1 PG-50), 50. https://dx.doi.org/10.1186/s12874-021-01244-7

John, A., Okolie, C., Eyles, E., Webb, R. T., Schmidt, L., McGuiness, L. A., Olorisade, B. K., Arensman, E., Hawton, K., Kapur, N., Moran, P., O'Connor, R. C., O'Neill, S., Higgins, J. P. T., & Gunnell, D. (2020). The impact of the COVID-19 pandemic on self-harm and suicidal behaviour: A living systematic review. *F1000Research*, *9*, 1097. https://doi.org/10.12688/f1000research.25522.1

Juul, S., Klingenberg, S., Gluud, C., Nielsen, N., Bentzer, P., Linder, A., Veroniki, A. A., Thabane, L., & Jakobsen, J. C. (2020). Interventions for treatment of COVID-19: A protocol for a living systematic review with network meta-analysis including individual patient data (The LIVING Project). *Systematic Reviews*, *9*(1 PG-108), 108. http://dx.doi.org/10.1186/s13643-020-01371-0

Kambhampati, S. B. S., Vaishya, R., & Vaish, A. (2020). Unprecedented surge in publications related to COVID-19 in the first three months of pandemic: A bibliometric analytic report. *Journal of Clinical Orthopaedics and Trauma*, *11*, S304–S306. https://doi.org/10.1016/j.jcot.2020.04.030

Khamis, A. M., Kahale, L. A., Pardo-Hernandez, H., Schunemann, H. J., & Akl, E. A. (2019). Methods of conduct and reporting of living systematic reviews: A protocol for a living methodological survey. *F1000Research*, *8*(PG-221), 221. https://dx.doi.org/10.12688/f1000research.18005.2

Kingdon, J. W. (2011). *Agendas, Alternatives, and Public Policies* (Longman Classics Edition, Ed.; Second Edi). HarperCollins College Publishers.

Korang, S. K., Juul, S., Nielsen, E. E., Feinberg, J., Siddiqui, F., Ong, G., Klingenberg, S., Gluud, C., Veroniki, A. A., Bu, F., Thabane, L., Thomsen, A. R., & Jakobsen, J. C. (2020). Vaccines to prevent COVID-19: A protocol for a living systematic review with network meta-analysis including individual patient data (The LIVING VACCINE Project). *Systematic Reviews*, *9*(1 PG-262), 262. http://dx.doi.org/10.1186/s13643-020-01516-1

Lansky, A., & Wethington, H. R. (2020). Living Systematic Reviews and Other Approaches for Updating Evidence. *American Journal of Public Health*, *110*(11 PG-1687–1688), 1687–1688. https://dx.doi.org/10.2105/AJPH.2020.305920

Lavis, J. N. (2021). *COVID-END taxonomy of public-health measures, clinical management of COVID-19, health-system arrangements, and economic and social responses*. https://www.mcmasterforum.org/docs/default-source/resources/resources_covid-taxonomy.pdf?sfvrsn=2b288648_5

Lavis, J. N., Oxman, A. D., Souza, N. M., Lewin, S., Gruen, R. L., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 9: Assessing the applicability of the findings of a systematic review. *Health Research Policy and Systems*, *7*(Suppl 1), S9. https://doi.org/10.1186/1478-4505-7-S1-S9

Lee, C., Thomas, M., Ejaredar, M., Kassam, A., Whittle, S. L., Buchbinder, R., Tugwell, P., Wells, G., Pardo, J. P., & Hazlewood, G. S. (2022). Crowdsourcing trainees in a living systematic review provided valuable experiential learning opportunities: A mixed-methods study. *Journal of Clinical Epidemiology*, *147*, 142–150. https://doi.org/10.1016/j.jclinepi.2022.03.019

Lerner, I., Atal, I., Crequit, P., & Ravaud, P. (2019). Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. *Journal of Clinical Epidemiology*, *108*(PG-86-94), 86–94. http://dx.doi.org/10.1016/j.jclinepi.2018.12.001

LIvE Framework. (2020). *Living Interactive Systematic Reviews*. https://rcc.network-meta-analysis.com/#

Macdonald, H., Loder, E., & Abbasi, K. (2020). Living systematic reviews at The BMJ. *BMJ (Clinical Research Ed.), 370*(PG-m2925), m2925. https://dx.doi.org/10.1136/bmj.m2925

Maguire, B. J., & Guerin, P. J. (2020). A living systematic review protocol for COVID-19 clinical trial registrations. *Wellcome Open Research*, *5*(PG-60), 60. https://dx.doi.org/10.12688/wellcomeopenres.15821.1

Martínez García, L., Pardo-Hernandez, H., Superchi, C., Niño de Guzman, E., Ballesteros, M., Ibargoyen Roteta, N., McFarlane, E., Posso, M., Roqué i Figuls, M., Rotaeche del Campo, R., Sanabria, A. J., Selva, A., Solà, I., Vernooij, R. W. M., & Alonso-Coello, P.

(2017). Methodological systematic review identifies major limitations in prioritization processes for updating. *Journal of Clinical Epidemiology*, *86*, 11–24. https://doi.org/10.1016/j.jclinepi.2017.05.008

Mavergames, C., & Elliott, J. (n.d.). *Living Systematic Reviews: Towards real-time evidence for health-care decision-making*. BMJ Best Practice. Retrieved July 12, 2022, from https://bestpractice.bmj.com/info/us/toolkit/discuss-ebm/living-systematic-reviews-towards-real-time-evidence-for-health-care-decision-making/

Moher, D., & Tsertsvadze, A. (2006). Systematic reviews: When is an update an update? *The Lancet*, *367*(9514), 881–883. https://doi.org/10.1016/S0140-6736(06)68358-X

Moher, D., Tsertsvadze, A., Tricco, A. C., Eccles, M., Grimshaw, J., Sampson, M., & Barrowman, N. (2007). A systematic review identified few methods and strategies describing when and how to update systematic reviews. *Journal of Clinical Epidemiology*, *60*(11), 1095.e1-1095.e11. https://doi.org/10.1016/j.jclinepi.2007.03.008

Moher, D., Tsertsvadze, A., Tricco, A., Eccles, M., Grimshaw, J., Sampson, M., & Barrowman, N. (2008). When and how to update systematic reviews. *Cochrane Database of Systematic Reviews*, *2010*(1). https://doi.org/10.1002/14651858.MR000023.pub3

Mondello, S., Sorinola, A., Czeiter, E., Vamos, Z., Amrein, K., Synnot, A., Donoghue, E., Sandor, J., Wang, K. K. W., Diaz-Arrastia, R., Steyerberg, E. W., Menon, D. K., Maas, A. I. R., & Buki, A. (2021). Blood-Based Protein Biomarkers for the Management of Traumatic Brain Injuries in Adults Presenting to Emergency Departments with Mild Brain Injury: A Living Systematic Review and Meta-Analysis. *Journal of Neurotrauma*, *38*(8 PG-1086–1106), 1086–1106. https://dx.doi.org/10.1089/neu.2017.5182

Nikolakopoulou, A., Egger, M., Salanti, G., Mavridis, Di., Furukawa, T. A., Cipriani, A., Tricco, A. C., Straus, S. E., & Siontis, G. C. M. (2018). Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: Empirical study. *BMJ (Online)*, *360*(PG-k585), k585. http://dx.doi.org/10.1136/bmj.k585

Nikolakopoulou, A., Mavridis, D., Egger, M., & Salanti, G. (2018). Continuously updated network meta-analysis and statistical monitoring for timely decision-making. *Statistical Methods in Medical Research*, *27*(5), 1312–1330. https://doi.org/10.1177/0962280216659896

Pearson, H. (2021). How COVID broke the evidence pipeline. *Nature*, *593*(7858), 182–185. https://doi.org/10.1038/d41586-021-01246-x

Rahal, A. K., Badgett, R. G., & Hoffman, R. M. (2016). Screening Coverage Needed to Reduce Mortality from Prostate Cancer: A Living Systematic Review. *PloS One*, *11*(4 PG-e0153417), e0153417. https://dx.doi.org/10.1371/journal.pone.0153417

Riaz, I. Bin, He, H., Ryu, A. J., Siddiqi, R., Naqvi, S. A. A., Yao, Y., Husnain, M., Narasimhulu, D. M., Mathew, J., Sipra, Q. U. A. R., Vandvik, P. O., Joseph, R. W., Liu, H., Wang, Z., Herasevich, V., Singh, P., Hussain, S. A., Ho, T. H., Bryce, A. H., … Costello, B. A. (2021). A Living, Interactive Systematic Review and Network Meta-analysis of First-line Treatment of Metastatic Renal Cell Carcinoma. *European Urology*, *PG-*. https://dx.doi.org/10.1016/j.eururo.2021.03.016

Riaz, I. Bin, Rawal, S. C., Siddiqi, R., Asghar, N., Akhtar, M., Gajic, O., Wang, Z., Montori, V., Herasevich, V., Go, R. S., Baral, C., Rajkumar, S. V., & Murad, M. H. (2019). Innovation in evidence synthesis: A living systematic review of immune checkpoint

inhibitors in cancer patients. *Journal of Global Oncology*, *5*(Supplement PG-80), 80. http://dx.doi.org/10.1200/JGO.2019.5.suppl.80

Riaz, I. Bin, Siddiqi, R., Asghar, N., Cathcart-Rake, E. J., Herasevich, V., Montori, V., Wang, Z., Go, R. S., Rajkumar, S. V., & Murad, M. H. (2019). Living systematic reviews: A novel mechanism for improving efficiency and quality of evidence synthesis in oncology. *Journal of Clinical Oncology*, *37*(27_suppl), 241–241. https://doi.org/10.1200/JCO.2019.37.27_suppl.241

Richard, M. E., Harricharan, S., & Anna, F. (2020). PCN291 Always up-to-Date Systematic Literature Reviews (SLRS) NOW a Reality—An Example in NEWLY Diagnosed ACUTE Myeloid Leukemia (ND AML) and Relapsed/Refractory ACUTE Lymphoblastic Leukemia (RR ALL). *Value in Health*, *23*, S473. https://doi.org/10.1016/j.jval.2020.08.428

Ritch, R. (2016). Living Reviews-A New Frontier for the Asia-Pacific Journal of Ophthalmology. *Asia-Pacific Journal of Ophthalmology (Philadelphia, Pa.)*, *4*(2 PG-64–5), 64–65. https://dx.doi.org/10.1097/APO.0000000000000114

Santillan-Garcia, A. (2020). Living evidence for SARS-CoV-2. *Medicina Intensiva*, *PG-*. http://dx.doi.org/10.1016/j.medin.2020.04.012

Santillan-Garcia, A., Bravo-Jeria, R., Verdugo-Paiva, F., & Rada, G. (2020). [Living evidence in response to controversies about the use of antimalarials in COVID-19]. *Evidencia Viva Como Respuesta a Las Controversias En El Uso de Antimalaricos En COVID-19.*, *73*(8 PG-693–694), 693–694. https://dx.doi.org/10.1016/j.recesp.2020.05.001

Shanahan, D. R. (2015). A living document: Reincarnating the research article. *Trials*, *16*(1), 151. https://doi.org/10.1186/s13063-015-0666-5

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*(1), 10. https://doi.org/10.1186/1471-2288-7-10

Shekelle, P. G., Shetty, K., Newberry, S., Maglione, M., & Motala, A. (2017). Machine Learning Versus Standard Techniques for Updating Searches for Systematic Reviews: A Diagnostic Accuracy Study. *Annals of Internal Medicine*, *167*(3), 213. https://doi.org/10.7326/L17-0124

Shokraneh, F., & Russell-Rose, T. (2020). Lessons from COVID-19 to future evidence synthesis efforts: First living search strategy and out of date scientific publishing and indexing industry. *Journal of Clinical Epidemiology*, *123*(PG-171-173), 171–173. http://dx.doi.org/10.1016/j.jclinepi.2020.04.014

Siemieniuk, R. A., Bartoszko, J. J., Ge, L., Zeraatkar, D., Izcovich, A., Kum, E., Pardo-Hernandez, H., Rochwerg, B., Lamontagne, F., Han, M. A., Liu, Q., Agarwal, A., Agoritsas, T., Chu, D. K., Couban, R., Darzi, A., Devji, T., Fang, B., Fang, C., … Cusano, E. (2020). Drug treatments for covid-19: Living systematic review and network meta-analysis. *BMJ (Clinical Research Ed.)*, *370*(PG-m2980), m2980. https://dx.doi.org/10.1136/bmj.m2980

Siqueira, I., Clark, E., Neil-Sztramko, S., Belita, E., & Dobbins, M. (2021). Informing policy on school and daycare operations during COVID-19 with a living rapid evidence review. *European Journal of Public Health*, *31*(Supplement_3). https://doi.org/10.1093/eurpub/ckab164.593

Spurling, G. K. P., Del Mar, C. B., Dooley, L., Clark, J., & Askew, D. A. (2017). Delayed antibiotic prescriptions for respiratory infections. *Cochrane Database of Systematic Reviews*, *9 PG-*. https://doi.org/10.1002/14651858.CD004417.pub5

Sutton, A. J., Donegan, S., Takwoingi, Y., Garner, P., Gamble, C., & Donald, A. (2009). An encouraging assessment of methods to inform priorities for updating systematic reviews. *Journal of Clinical Epidemiology*, *62*(3), 241–251. https://doi.org/10.1016/j.jclinepi.2008.04.005

Takwoingi, Y., Hopewell, S., Tovey, D., & Sutton, A. J. (2013). A multicomponent decision tool for prioritising the updating of systematic reviews. *BMJ*, *347*(dec13 1), f7191–f7191. https://doi.org/10.1136/bmj.f7191

The Cochrane Collaboration. (n.d.). *Living systematic reviews*. Retrieved July 12, 2022, from https://community.cochrane.org/review-production/production-resources/living-systematic-reviews

Turner, T., Elliott, J., Tendal, B., Vogel, J. P., Norris, S., Tate, R., & Green, S. (2022). The Australian living guidelines for the clinical care of people with COVID-19: What worked, what didn't and why, a mixed methods process evaluation. *PLOS ONE*, *17*(1), e0261479. https://doi.org/10.1371/journal.pone.0261479

Vandvik, P. O., Brignardello-Petersen, R., & Guyatt, G. H. (2016). Living cumulative network meta-analysis to reduce waste in research: A paradigmatic shift for systematic reviews? *BMC Medicine*, *14*(PG-59), 59. https://dx.doi.org/10.1186/s12916-016-0596-4

Vergara-Merino, L., Verdejo, C., Carrasco, C., & Vargas-Peirano, M. (2020). Living systematic review: New inputs and challenges. *La Revision Sistematica Viva: Nuevos Aportes y Desafios.*, *20*(11 PG-e8092), e8092. https://dx.doi.org/10.5867/medwave.2020.11.8092

Vickery, J., Atkinson, P., Lin, L., Rubin, O., Upshur, R., Yeoh, E.-K., Boyer, C., & Errett, N. A. (2022). Challenges to evidence-informed decision-making in the context of pandemics: Qualitative study of COVID-19 policy advisor perspectives. *BMJ Global Health*, *7*(4), e008268. https://doi.org/10.1136/BMJGH-2021-008268

Vogel, J. P., Dowswell, T., Lewin, S., Bonet, M., Hampson, L., Kellie, F., Portela, A., Bucagu, M., Norris, S. L., Neilson, J., Guelmezoglu, A. M., & Oladapo, O. T. (2019). Developing and applying a "living guidelines" approach to WHO recommendations on maternal and perinatal health. *BMJ Global Health*, *4*(4 PG-). https://doi.org/10.1136/bmjgh-2019-001683

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., Schmid, C. H., Bertram, L., Lill, C. M., Cohen, J. T., & Trikalinos, T. A. (2012). Toward modernizing the systematic review pipeline in genetics: Efficient updating via data mining. *Genetics in Medicine*, *14*(7), 663–669. https://doi.org/10.1038/gim.2012.7

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., Aravkin, A. Y., Bisignano, C., Barber, R. M., Alam, T., Fuller, J. E., May, E. A., Jones, D. P., Frisch, M. E., Abbafati, C., Adolph, C., Allorant, A., Amlag, J. O., Bang-Jensen, B., … Murray, C. J. L. (2022). Estimating excess mortality due to the COVID-19 pandemic: A systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, *399*(10334), 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3

Winters, M., de Vos, R.-J., van Middelkoop, M., Rathleff, M. S., & Weir, A. (2021). Stay alive! What are living systematic reviews and what are their advantages and challenges? *British Journal of Sports Medicine*, *55*(10 PG-). https://doi.org/10.1136/bjsports-2020-103490

4.11. Additional files

Additional file 1. Search strategies

Database	Search strategy
OVID (Medline and EMBASE)	(living or live).mp. adj 4 (reviews or review or evidence or synthes* or meta-analy* or metanaly* or "meta analysis" or "meta analyses" or map or maps or overview* or SR).mp
Cochrane	(living or live) NEAR/4 (reviews or review or evidence or synthes* or meta-analy* or metanaly* or "meta analysis" or "meta analyses" or map or maps or overview* or SR)
WebofScience (excludes Medline)	TS = ((living OR live) NEAR/4 (reviews OR review OR evidence OR synthes* OR meta-analy* OR metanaly* OR "meta analysis" OR "meta analyses" OR map OR maps OR overview* OR SR)))
PubMed	(living or live) n4 reviews or review or evidence or synthes* or meta-analy* or metanaly* or "meta analysis" or "meta analyses" or map or maps or overview* or SR)
HealthSystemsEvidence	(living OR live) AND (reviews OR review OR evidence OR synthesis OR syntheses OR meta-analysis OR metanalysis OR "meta analysis" OR

	"meta analyses" OR map OR maps OR overview OR SR)

Additional file 2. Excluded articles.

List of articles excluded specifying their reasons. Available upon request

Additional file 3. Methodological limitations (MMAT) of the empirical primary studies included.

Ref ID	**Title**	**1. QUALITATIVE**					**3. QUANTITATIVE NON-RANDOMIZED**					**4. QUANTITATIVE DESCRIPTIVE**					**5. MIXED METHODS**				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Millard	Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation																N	Y	N	N	Y
Gates	LOCATE: a prospective evaluation of the value of Leveraging Ongoing Citation Acquisition Techniques for living Evidence syntheses											Y	CT	Y	CT	Y					
Kingdon	Agendas, Alternatives, and Public Policies	Y	Y	Y	Y	CT															
Arienti 2022	The methodology of a "living" COVID-19 registry development in a clinical context.	Y	N	Y	N	CT															
Pierre 2022	Secondary electronic sources demonstrated very good sensitivity for identifying studies evaluating interventions for COVID-19											Y	Y	Y	Y	Y					

Lee 2022	Crowdsourcing trainees in a living systematic review provided valuable experiential learning opportunities: a mixed-methods study															Y	Y	Y	N	N
Perlman-Arrow 2022	A real-world evaluation of the implementation of NLP technology in abstract screening of a systematic review											N	Y	Y	C T	Y					
Metzendorf 2021	Evaluation of the comprehensiveness, accuracy, and currency of the Cochrane COVID-19 Study Register for supporting rapid evidence synthesis production											Y	N	Y	Y	Y					
Turner 2022	The Australian living guidelines for the clinical care of people with COVID-19: What worked, what didn't and why, a mixed methods process evaluation											N	Y	Y	N	N					
Turner 2022	How frequently should "living" guidelines be updated? Insights from the											C T	Y	Y	N	Y					

	Australian Living Stroke Guidelines																			

CT: Can’t tell; N: No; Y: Yes

Additional file 4. Contribution of each article to the thematic categories

Details on how each paper included contributes to each thematic category. Available upon request.

Additional file 5. Detailed description of each thematic category

Additional table 5.1. Thematic category 1: definition of a living evidence synthesis

	Sub-theme 1. Living evidence syntheses	**Sub-theme 2. Updates**	**Sub-theme 3. Labelling 'living'**
Short description	Understanding what constitutes a living synthesis	Understanding what constitutes an 'update' in the context of an LES?	What do we understand by the label 'living' in the context of evidence syntheses
Details	An LES understood as a summary of research that: • summarizes all the existing evidence in a cumulative way; • at any defined point in time is up-to-date (i.e., it is continually updated as evidence becomes available); • can address single comparisons or multiple available comparisons (i.e., living network meta-analysis); • [similar to a non-living evidence synthesis] can use different type of methods to combine studies (narrative summary, meta-analysis, network meta-analysis); • can use different channels to be disseminated, but it would most likely have an online up-to-date summary available.	An update can be understood as any change in an evidence synthesis that: • adds new evidence that was not previously included, coming from: ○ new research that fits eligibility criteria that was made available since the last search date; ○ new research that adds new interventions/comparisons for a given condition (i.e., in a living network meta-analysis); ○ research that was available before the last search date, but it was not included in the original document because of: ▪ limitation in the search strategy/screening process; ▪ research was produced before, but was made available after the search date (e.g., historical material released at a specific time); • changes the eligibility criteria (and the protocol); • updates search strategies (i.e., living search strategy);	Different approaches to label as 'living' a given evidence synthesis, including: • understanding the label 'living' as a transient status (i.e., one synthesis could be living only for a specific period of time, and could stop being living at some point); • understanding the label 'living' as a scale rather than a status (i.e., 'livingness' of a synthesis); • understanding the label 'living' as a *a priori* commitment from synthesisers (i.e., including a clear plan on how to incorporate new evidence when it becomes available and a credible commitment about when updates will be available). Considering whether it is easier to label an evidence synthesis as non-living or developing criteria for an evidence synthesis to be out-of-date

		• changes the presentation details of the synthesis (e.g., living document). In studies addressing what should be considered an update in the context of non-living evidence syntheses, there is a contested definition of what should be considered an update (i.e., whether only adding new evidence should be considered a real update).	

Additional table 5.2. Thematic category 2: methods to assess the need and to produce a living evidence synthesis.

	Sub-theme 1. Assessment of the need of a living evidence synthesis (or to update one)	**Sub-theme 2. Team management**	**Sub-theme 3. Production**
Short description	Methods to predict whether new literature might change the findings.	Methods to facilitate the team management while producing a living evidence synthesis	Methods to facilitate the production of a living evidence synthesis
Details	Two groups of (non-exclusive) methods that assess: • the probability that new literature might change the existing findings (e.g., Ottawa, RAND, etc.) or the value that new information could provide to reduce uncertainty. • how susceptible the existing findings are to be changed if new evidence arises (e.g., GRADE certainty of the evidence is low or very low for a given outcome). • how the context or the issue might change the applicability of the findings or the understanding of a giving phenomena.	Strategies to facilitate the work in teams: • Streamline pathways, workflows, and role definitions, by setting individual small tasks that facilitate a manageable workload across time. • Outlining clear boundaries where technology can facilitate specific tasks. Strategies to ensure team sizes that would make feasible the production of a living evidence synthesis: • Consider the type of incentives that authors might get to participate (e.g., authorship).	Methods to facilitate searching and study selection: • Artificial intelligence (including machine learning, or natural language processing to check if new citations are not already included), data mining and neural networks to facilitate periodic searches (which could also be complemented with semi-automated approaches) • Living collections of evidence (curated libraries or repositories for specific topics) that can include pre-print servers. • Alerts from some bibliographic databases

		• Having topic-oriented living evidence synthesis communities that would be 'ready' to take on new syntheses. • Crowdsourcing • Use of trainees	• Librarian help desks to conduct and adapt search strategies (mediated search services) Methods to facilitate data extraction and risk-of-bias assessment: • Artificial intelligence and machine learning (which could also be complemented with semi-automated approaches) • Linked data from studies that have already been processed as part of another synthesis. Methods to facilitate data synthesis: • Artificial intelligence and machine learning (which could also be complemented with semi-automated approaches) • Statistical approaches to manage the increase of type I error in repeated meta-analyses (e.g., trial sequential analysis, sequential meta-analysis, Shuster method, Bayesian frameworks, etc.) • Text templates for synthesizing evidence.

Additional table 5.3. Thematic category 3: when to produce a living evidence synthesis.

	Sub-theme 1. Type of decisions	Sub-theme 2. Demand side triggers	Sub-theme 3. Supply side triggers	Sub-theme 4. Other elements to consider
Short description	Alternatives for an evidence producer when starting a new evidence synthesis.	'Triggers' when a living evidence synthesis might need to be produced		Other elements to be considered when making a decision of when a living synthesis needs to be conducted
Details	At any given time, an evidence synthesis producer has the following options: • starting an evidence synthesis (regardless of whether it will become living or not) – i.e., 'baseline' synthesis; • updating an existing evidence synthesis as a one-off exercise; • updating an existing evidence synthesis in a living way: o making a non-living evidence synthesis (most commonly the 'baseline' synthesis) living (which is the focus of the next table); o keeping living an existing living evidence synthesis (which is also the focus of the next table).	Demand-side triggers are determined when a topic is relevant for decision making, which could be influenced by: • agenda-setting dynamics (coupling of problems, policies, and politics streams), including the role of stakeholders that are aiming to influence decision maker agendas; • how the context and issues might influence agenda-setting dynamics (e.g., new variants emerge). • urgency of a given decision (i.e., immediate answer to pressing needs); • other factors that influence priorities in research priority setting processes (e.g., supplier induced demand, social media); • a need for having 'living' recommendations (or other evidence-informed products	Supply-side triggers could be coming from two bodies of evidence: • existing body of evidence: susceptibility of the current findings to change if more evidence is added (i.e., certainty of the existing body of evidence, or level of saturation of a given concept, or immatureness of the existing evidence base) • New non-synthesized evidence is available and could change the current findings because: o It will change the principal findings or summary estimates (e.g., effect estimates) o It will change the certainty or confidence in the existing evidence (e.g., increasing precision of pooled estimates, or the coherence of the findings of multiple studies)	Other elements might need to be considered by researchers when making decisions about starting a living evidence synthesis: • workload and availability of the research team. • time that usually takes for a primary study to be included in an evidence synthesis; • the question has been already addressed by a different evidence synthesis; • new developing methods would allow a different analysis of the same data; • critical feedback received from evidence synthesis readers; • presence of publication bias, indicating the need to update the synthesis to reduce this potential effect (the risk of publication bias might be amplified too, since trials with positive results may be

		to promote the use of evidence in decision making e.g., living knowledge translation).	○ It will solve a 'lingering controversy' for a given topic. ○ It might change the relevance of the existing question or issue (e.g., new political power balance, new technology that changes an acceptable control group, etc.)	published sooner than trials with negative results).

Additional table 5.4. Thematic category 4: when to update a living evidence synthesis (keep a living evidence synthesis 'living')

	Sub-theme 1. Processed involved	Sub-theme 2. Triggers from the demand side	Sub-theme 3. Triggers from the supply side	Sub-theme 4. Frequency
Short description	Parts of an evidence synthesis that could be updated.	'Triggers' when a living evidence synthesis might need to be updated		Deciding how frequent a living evidence synthesis needs to be updated.
Details	Depending on the demand and supply side 'triggers' (including the role of funders or intermediaries), for a given update to a living evidence synthesis, the following parts of the synthesis could be updated: • eligibility criteria. • search strategies ('adaptive' search strategies); • data synthesis; • data appraisal; • publication.	A potential change in any of the 'triggers' outlined in the previous table. Common issues to consider might be: • whether or not the research question is still a priority for decision making (e.g., whether the first version of the synthesis continues to be used frequently); • an existing synthesis has already addressed the topic.	Similar to the previous table, supply-side triggers could be coming from two bodies of evidence: • Susceptibility of the findings from an existing body of evidence to change if more evidence is added (i.e., certainty of the existing body of evidence, or level of saturation of a given concept) • New non-synthesized evidence is frequently made available (i.e., topic remains active from a research perspective; e.g., large number of studies underway, type of policy being implemented in several places and could be evaluated), and could change the current findings because: o It will change the summary estimates (e.g., effect estimates)	Two main approaches could be selected to decide how frequently a living evidence synthesis will get updated: • regular frequency (e.g., weekly, monthly, quarterly, etc.); • tailored frequency (that could also change over time), based on the occurrence of 'triggers' outlined in subthemes 2 and 3, and the negotiations with the demand. A number of supportive tools can help researchers decide on defining either approach: • methods to estimate the needed frequency are outlined in a previous table (could be done by setting a specific number or scale of new studies available to trigger a new update);

			○ It will change the certainty of the existing body of evidence. ○ It will solve a 'lingering controversy' in a given topic.	• having a steering committee that can assess the 'triggers' and can advise researchers on what part of an evidence synthesis should be updated (sub-theme 1). Regardless of the alternative chosen by researchers, researchers should engage in credible commitments so the frequency being communicated is the actual frequency that a living evidence synthesis is updated.

Additional table 5.5. Thematic category 5: dissemination of the findings of a living evidence synthesis

	Sub-theme 1. Platforms	**Sub-theme 2. Structured format**	**Sub-theme 3. Living evidence syntheses users**	**Sub-theme 4. Speeding-up**
Short description	Platforms that could be used to make available the findings	Adaptations to the format that can be used to streamline dissemination processes	Types of decision-makers and evidence intermediaries that can use the findings of LESs	Strategies that can be used to reduce the time from which findings are available and are used by decision-makers and evidence intermediaries.
Details	Different platforms (and combinations) to make available the findings of a living evidence synthesis are available: • website; • scientific journal; • interactive platforms (e.g., apps) and infographics. These platforms respond differently to the multiple challenges and opportunities that a living evidence synthesis creates. Some of these challenges and opportunities are: • Limited number of words and tables/figures to present the findings. • Presenting results in alternative formats (e.g., videos)	A number of different parts within a living evidence synthesis could be accommodated to allow the regular updating of evidence, and allowing decision-makers and evidence intermediaries to easily find what they were looking on a given context: • creating a section in each update that highlights the changes from the previous version; • providing clear information for users on what parts of the synthesis were updated (e.g., number of new studies added using, for example, PRISMA standards, search dates, etc.), even though the findings do not change);	The following types of users might be using the findings of a living evidence synthesis: • primary researchers (e.g., trialists might want to see their findings in the context of other similar studies); • intermediate users (guidelines, HTA recommendations) • decision makers. Evidence producers might want to adapt their LES findings to improve the user experience and thereby better support their uptake	A number of strategies could be used to reduce the time in which findings are available for decision-makers and evidence intermediaries (i.e., the latest version is available as quickly as possible): • Pre-print servers to publish the non-peer synthesised results in advance; • for living evidence syntheses published in a scientific journal: o abbreviated submissions of updates (e.g., only the findings tables and abstracts were updated); o short commentary published to flag that searches were updated; o previous updates are kept as an appendix; o introducing flexibility of the authorship criteria (e.g., IMJE) for

	• Length of the editorial process (including peer-synthesis) • Indirect benefits associated with the reputation of publishing an article in a high-impact scientific journal. • Authorship of each update • Availability of previous updates (e.g., in an appendix, different DOI, different URL, etc.) • Possibility to improve user experience with the data synthesized (e.g., interactive summary of findings tables) • Further interaction with users, so they can contribute on improving the living evidence synthesis in future updates.	• making explicit credible commitments about the frequency to expect a new update; • detailing a clear and explicit and transparent update plan.		allowing 'evolving authorship'; ○ identifying a pool of researchers by topic= available to act as peer synthesisers in a fast way; • notifying readers that the searches were updated, even though there were no changes to the findings.

Additional table 5.6. Thematic category 6: when to discontinue updates of a living evidence synthesis.

	Sub-theme 1. Demand-side triggers	Sub-theme 2. Supply-side triggers	Sub-theme 3. Other elements to consider
Short description	'Triggers' when a living evidence synthesis could stop being updated		Other elements to be considered when making a decision of when a living synthesis can stop being updated.
Details	A potential change in any of the 'triggers' outlined in previous tables. Common issues to consider might be: • whether or not the research question is still a priority for decision making or if it can be re-framed; • the issue might not be a priority anymore, but it could have strategic importance (e.g., pandemic preparedness).	A potential change in any of the 'triggers' outlined in previous thematic categories. Common issues to consider might be: • the findings using the latest body of evidence are unlikely to change (e.g., high certainty of the evidence, reached saturation); • no new evidence is expected to be available.	Other elements that might be considered are: • engagement of the synthesis team with the process and final product (e.g., website, journal publication, etc.); • planned obsolescence of a funded project; • contextual variables, (e.g., alternative solutions to address the same issue are available, availability of vaccines made natural immunity questions less important).

Chapter 5: Conclusions

This thesis includes three studies that address two main issues for evidence-support systems. Studies 1 and 2 (included in chapters 2 and 3) contribute to better aligning evidence demand (what decision-makers need) with evidence supply (evidence support that draws on existing evidence or, when time allows, new primary research can offer), by facilitating the process of mapping a decision-makers' question to the right study design to address their question. Study 3 (included in chapter 4) provides a conceptual framework to better understand the role of living evidence syntheses (LESs) in decision-making processes. Together, the three studies contribute to strengthen evidence-support systems and the global evidence architecture by better aligning evidence demand and supply, as well as providing insights to better understand the role of LESs in this context.

This chapter presents key findings of the thesis, highlighting substantive, theoretical and methodological contributions, followed by what the thesis adds to the current literature. The chapter ends with a summary of the strengths and limitations of the work, and implications for practice and policy and for future research.

5.1. Summary of findings and contributions

This thesis provides important contributions to evidence-support systems and evidence-informed decision-making. Contributions are separated into substantive, theoretical and methodological ones. The main substantive contribution is to highlight two key issues of critical importance to

evidence-support systems, and that might be crucial to their growth and evolution in the near future.

The first issue is the connection between evidence demand and supply. This thesis creates a practical tool that facilitates the alignment between evidence demand and supply, by receiving a decision-maker question as an input and providing a suitable study design to answer that question as an output. This tool includes two critical steps: 1) mapping the question to one in a newly developed taxonomy; and 2) matching that question to a suitable study design or a list of potential study designs.

The second issue that is addressed in this thesis is the role of LESs in decision-making processes, which is advanced by producing a framework to understand what are, and when to produce (update, and stop update) LESs for decision-making purposes. As LESs are becoming a critical component of evidence-support systems, in part spurred by their primacy during the COVID-19 pandemic (due to their capacity to evolve as the context, issues and evidence evolved), this issue is likely to play a critical role in the near future and for future global crises.

Secondly, this thesis makes one principal theoretical contribution by providing a comprehensive framework for the role of LESs in decision-making. Chapter 4 presents the results of a critical interpretive synthesis (CIS), with six emerging themes from the literature: (1) What are LESs (and their updates)?; (2) What methodological support exists for producing LES?; (3) When to produce an LESs?; (4) When to update an LES?; (5) How to make available the findings of an LES?; an (6) When discontinue updating an LES? These themes were included in a conceptual

framework that includes a visual representation, describing demand-side (from decision-makers) and supply-side (from evidence intermediaries and producers) triggers to produce, update or stop updating an LES.

Finally, this thesis provides a methodological contribution in generating a new method or tool that facilitates the connection between a decision maker need and the right study design to address that need. Decision-makers can now ask a specific question, which can be mapped against the taxonomy of types of question explained in chapters 2, and then in turn mapped the right study design to address that question (which is explained in chapter 3).

Chapter 2 creates a mutually exclusive and collectively exhaustive list of questions for which decision-makers could expect evidence to provide insights. This study was conducted using a demand-driven approach by asking evidence-support units to provide the list of questions that they have answered and then creating a taxonomy of types of questions. Forty different types of question were identified and organized in goals to be achieved, and the goals were grouped in the four different decision-making stages (clarifying a societal problem, its causes, and potential impacts; finding and selecting options to address a problem; implementing or scaling-up an option; and monitoring implementation and evaluating impacts).

Chapter 3 builds on the findings of study 1 (presented in chapter 2) and creates a list of study designs most suitable to answer each one of the 40 different types of question identified in study 1. To do so, an online two-round Delphi survey was used to reach a consensus across each one of the 40 different types of questions. A consensus could be reached for the majority of the types of

question included (28 types of question), and a ranking of study design for these types of question is presented in Chapter 4.

5.2. Findings in relation to the existing literature

As mentioned in the previous section, this thesis builds on the existing literature in two key areas for evidence-support systems, namely facilitating the alignment between evidence demand and supply (by clearly mapping different types of question and the study designs that better address them) and the role of LESs in the decision-making process (by providing a conceptual framework for what are and when to produce and update LESs). Together, the three studies contribute to move evidence-support systems and global evidence architecture to the next level and better aligning evidence demand and supply with a pivotal role of LESs.

The connection of decision-maker needs with the type of study design is a critical issue to close the gap between evidence production and its uses for decision-making. While multiple barriers to using evidence to make decisions have been identified (Oliver et al., 2014), the interface between evidence demand and supply is a critical place to look when ensuring that decisions are informed with the best available evidence.

On the other hand, while the traditional hierarchy of evidence was a useful tool to orient and train researchers on what type of study designs were more robust than others, it has now been contested by multiple actors, mainly because the study design used might differ depending on the type of question that a study is trying to answer (Hansen, 2014; Petticrew, 2003). Similarly, more recent guidance has accepted that observational studies might provide high quality evidence for

prognostic studies (Foroutan et al., 2020) and that they could have some potential to contribute to pool estimates (Cuello et al., 2022), depending on the specific type of question that a research study is trying to address.

This thesis goes one step forward by creating a full list of 40 different types of question, and a list (and some cases hierarchies) of study designs that most suitably answer these questions. Together, these two chapters (chapters 2 and 3) provide concrete guidance on deciding what study design needs to be used based on the type of question that was originally formulated.

While other efforts have been made to explore the relationship between different study designs depending on the type of question asked (Evans, 2003; Hansen, 2014), none of them have used the demand-driven approach used in this thesis. The fact that the types of question found here were coming from questions that were actually asked by a decision-maker makes the taxonomy stronger. Moreover, previous efforts have not always considered the use of multiple methodologies (i.e., quantitative, and qualitative). In this thesis, both quantitative and qualitative methodologies were considered as potential options to address the types of question that were classified.

In terms of LESs, they were a common literature subject after the original paper series in 2017 (Elliott et al., 2017). While the field evolved, and initiatives such as the Cochrane Collaboration created a dedicated group to promote the production of Cochrane LESs (Cochrane Collaboration, 2017), it was the COVID-19 pandemic that brought LESs into the spotlight. Important initiatives leveraged the role that LESs should have in addressing the COVID-19 pandemic and future

global crises (*Collaborating in Response to COVID-19*, 2020), making LESs a key element to help evidence ecosystems deal with the unique challenges that the COVID-19 pandemic brought (Pearson, 2021).

In this context, Chapter 4 provides a unique framework bringing together multiple insights from the literature, to take this field one step ahead by questioning what LESs are, when we need to produce/update them, and how we can make their findings available. While other frameworks have been produced, they have mainly been focused on living guidelines (El Mikati et al., 2022), and have not used a CIS approach, which takes the existing literature as a source of themes to develop a comprehensive framework (Dixon-Woods et al., 2006).

5.3. Strengths and limitations

This thesis as a whole has many strengths. First, it addresses two critical issues that are at the center of the process to strengthen evidence-support systems. These issues are the alignment between evidence demand and supply (by facilitating a unique tool to connect decision makers' needs with study designs), and the role of LESs for decision making (by producing a conceptual framework of what are and when to produce and update LESs).

Secondly, it uses both a demand-driven and supply-driven approaches in separate parts of the thesis. On the one hand, a demand-driven approach is used in chapters 2 and 3 by asking evidence units to provide the list of questions that they have received from decision-makers, to create a taxonomy of types of question, and by asking methodologists to identify optimal study

designs for each question. On the other hand, it builds a conceptual framework for LESs based on the themes that emerge from the existing literature (supply side).

Thirdly, each chapter uses a variety of rigorous methods to produce strong and robust outputs (a novel tool to connect evidence demand and supply, and a conceptual framework for the role of LESs in decision-making). A global cross-sectional survey is used in chapter 2, an online ranking-type Delphi is used in chapter 3, and a critical interpretive synthesis in chapter 4.

Finally, this thesis is produced at a critical time, taking advantage of the momentum that has been created by the COVID-19 pandemic on building stronger domestic evidence-support systems and global evidence architecture to collectively think and ensure that decision-making is informed by the best available evidence.

This thesis also has limitations. First, the connection that is made between decision-makers' needs and the evidence support is conducted up to the level of study design only. A number of other evidence-support considerations might be also taken into account when making choices on how evidence could answer a specific decision-making need, such as data analysis methods, or the interaction across other forms of evidence (e.g., primary studies or evidence syntheses), but they are not addressed in this thesis.

Secondly, despite using a broader lens to incorporate insights from any sector, most of the participants of the studies included in chapters 2 and 3, and the literature consulted in chapter 4, are drawn from the health sector where much of the related activity has been happening.

5.4. Implications

5.4.1. Implications for policy and practice

This thesis has a number of implications for practice and policy that can be grouped into domestic (i.e., affecting only a given country or territory) and global implications.

For domestic-level implications, this thesis could facilitate the work of evidence intermediaries, decision-makers, evidence producers and research funders by facilitating their work.

Evidence intermediaries are entities that operate in the interface between evidence demand and supply and ensure that the advice provided to decision makers can be supported by the best available evidence. In this sense, this thesis might facilitate their work, by providing them with a mechanism to navigate from a decision-maker's need to finding a study design to address the specific formulated question. In this sense, an evidence intermediary could use this taxonomy to facilitate its work on translating a decision-makers' need into a clear and relevant research question, that could later be addressed by evidence producers or themselves. This might not always be a linear and easy process, but having this tool available would enable evidence intermediaries to go back and forth and clearly formulate an evidence need into a demand-driven type of research question.

It can also facilitate the work of decision-makers (demand-side) to help them ask better questions, in terms of what they can expect from evidence-support mechanisms. Hence, decision-makers would be able to easily scan the taxonomy produced in chapter 2 to understand and better formulate a specific question. Also, it can help evidence producers (supply-side), and

particularly researchers producing evidence synthesis, by helping them to coordinate in a demand-driven way, which can reduce duplication of efforts. Hence, the fact that decision-makers' needs could be mapped into a specific type of question and suggest a clear set of study designs narrows down the groups of evidence producers that have the adequate expertise to potentially provide an evidence-informed answer to that need. Finally, it can help entities funding evidence support and production to concentrate and better increase the value for money of their investments, by helping them to efficiently map evidence producers based on their capacity to address a specific decision-maker need.

We understand evidence-support systems as the set of institutions and structures that makes possible evidence-related advice for decision-making. In this context, this thesis helps to make evidence-support systems more robust. This could be facilitated by the outputs produced by studies 1 and 2 (i.e., by facilitating the translation between an evidence need and a clear study design that could robustly address that need). Additionally, a strong evidence-support system could also better coordinate on key areas where LESs might be needed, avoiding the un-coordinated production of multiple non-living low-quality evidence syntheses.

In the global-level implications, this thesis could contribute to better coordination within the global-evidence architecture. This is aligned with the 2023 implementation priorities established by the Evidence Commission (Global Commission on Evidence to Address Societal Challenges, 2023) on working with UN entities to better coordinate their work, alongside the better coordination of global evidence producers (such as the Cochrane and Campbell Collaborations) to make their work more demand driven.

On one side, this thesis might facilitate the work of multilateral organizations and global funders to help prioritize specific areas or questions in which global public goods (such as LESs) might be needed. On the other side, it might also help global evidence producers to facilitate their work, increasing the coordination level with the evidence demand, and reducing duplication among other groups that are producing evidence.

In terms of LESs, two critical implications can be derived from the outputs of this thesis. First, it might facilitate the work of evidence producers, by helping them to make better decisions on when to start, update, or stop updating a specific LES, by critically analyzing demand and supply triggers. Additionally, it might help evidence producers to decide on what ways they can make available the findings of the synthesis. Secondly, the conceptual framework (presented in chapter 4) might help decision-makers and evidence intermediaries to better understand what LESs are, and their role in the decision-making process.

5.4.2. Implications for future research

A number of implications for research emerge from this thesis. First, and the most important implication is that future research can empirically test the existing outputs of Chapters 2 and 3 (taxonomy and list of study designs) and framework (in Chapter 4) that were produced in this thesis, to see how they behave in concrete cases and broader audiences. For example, a series of qualitative focus groups with key stakeholders (which would include different types of decision makers, such as government policymakers and citizens) could be used to test whether the tool

created in Chapters 2 and 3 is actually helping to translate a decision-maker's need into a study design that can address these questions.

Similarly, a qualitative case study could test the conceptual framework created in Chapter 4, providing insights into the views of stakeholders regarding the role of LESs in decision-making, and whether the framework would be useful to make decisions on what key areas would be needed to be served by LESs, as opposed to a series of un-coordinated non-living syntheses.

Secondly, further research could find additional ways to operationalize the outputs of this thesis, such as by creating a tool to facilitate the application of the conceptual framework for LESs. For example, rounds of consultations with key stakeholders, or user testing could be conducted to explore the usability of this tool and framework.

Finally, future research could also help to disentangle how domestic evidence-support systems versus global evidence architecture might behave in their interaction with the outputs of this thesis. Hence, asking key stakeholders to explore whether decision-makers at a global or regional level ask different questions than local decision-makers, or whether global versus local evidence producers could create different impacts might be important to understand.

5.5. References

Cochrane Collaboration. (2017). *Living Evidence Network Governance Structure*. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/Living%20Evidence%20Network%20-%20Governance%20Structure_1.pdf

Cochrane Convenes. (2022). *Preparing for and responding to global health emergencies: Learnings from the COVID-19 evidence response and recommendations for the future. February 2022*. https://doi.org/10.6084/m9.figshare.19115849.v2

Collaborating in response to COVID-19: Editorial and methods initiatives across Cochrane. (2020). Wiley. https://doi.org/10.1002/14651858.CD202002

Cuello, C. A., Morgan, R. L., Brozek, J., Verbeek, J., Thayer, K., Ansari, M. T., Guyatt, G., & Schünemann, H. J. (2022). Case studies to explore the optimal use of randomized and nonrandomized studies in evidence syntheses that use GRADE. *Journal of Clinical Epidemiology*, *152*, 56–69. https://doi.org/10.1016/j.jclinepi.2022.09.014

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, *6*, 1–13. https://doi.org/10.1186/1471-2288-6-35

El Mikati, I. K., Khabsa, J., Harb, T., Khamis, M., Agarwal, A., Pardo-Hernandez, H., Farran, S., Khamis, A. M., El Zein, O., El-Khoury, R., Schünemann, H. J., Akl, E. A., Living Guidelines Group†, Alonso-Coello, P., Alper, B. S., Amer, Y. S., Arayssi, T., Barker, J. M., Bouakl, I., … Wiercioch, W. (2022). A Framework for the Development of Living

Practice Guidelines in Health Care. *Annals of Internal Medicine*, *175*(8), 1154–1160. https://doi.org/10.7326/M22-0514

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., … Pearson, L. (2017). Living systematic review: 1. Introduction—The why, what, when, and how. *Journal of Clinical Epidemiology*, *91*, 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84. https://doi.org/10.1046/j.1365-2702.2003.00662.x

Foroutan, F., Guyatt, G., Zuk, V., Vandvik, P. O., Alba, A. C., Mustafa, R., Vernooij, R., Arevalo-Rodriguez, I., Munn, Z., Roshanov, P., Riley, R., Schandelmaier, S., Kuijpers, T., Siemieniuk, R., Canelo-Aybar, C., Schunemann, H., & Iorio, A. (2020). GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in the identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology*, *121*, 62–70. https://doi.org/10.1016/j.jclinepi.2019.12.023

Global Commission on Evidence to Address Societal Challenges. (2023). *Evidence Commission update 2023: Strengthening domestic evidence-support systems, enhancing the global evidence architecture, and putting evidence at the centre of everyday life*. McMaster Health Forum. https://www.mcmasterforum.org/docs/default-source/evidence-commission/update-2023.pdf?sfvrsn=e81cbf_8

Hansen, H. F. (2014). Organisation of evidence-based knowledge production: Evidence hierarchies and evidence typologies. *Scandinavian Journal of Public Health*, *42*(13_suppl), 11–17. https://doi.org/10.1177/1403494813516715

Kuchenmüller, T., Lavis, J., Kheirandish, M., Reveiz, L., Reinap, M., Okeibunor, J., Siswanto, S., Rashidian, A., Sieber, S., Moat, K., Mansilla, C., El-Jardali, F., Helble, M., Reeder, J., Chapman, E., Barreto, J. O. M., Mandil, A., & Swaminathan, S. (2022). Time for a new global roadmap for supporting evidence into action. *PLOS Global Public Health*, *2*(6), e0000677. https://doi.org/10.1371/journal.pgph.0000677

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*, *14*(1), 2. https://doi.org/10.1186/1472-6963-14-2

Oxman, A., Lavis, J., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed health Policymaking (STP) 1: What is evidence-informed policymaking? *Health Research Policy and Systems*, *7*(Suppl 1), S1. https://doi.org/10.1186/1478-4505-7-S1-S1

Pearson, H. (2021). How COVID broke the evidence pipeline. *Nature*, *593*(7858), 182–185. https://doi.org/10.1038/d41586-021-01246-x

Petticrew, M. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527–529. https://doi.org/10.1136/jech.57.7.527