Chapter 4. Studies, syntheses and guidelines: Supply of evidence

4.1 Forms in which evidence is typically encountered in decision-making 44

4.2 Definitions of forms in which evidence is typically encountered 45

4.3 Matching decision-related questions to forms of evidence 47

4.4 Interplay of local and global evidence 48

4.5 Distinguishing high- from low-quality evidence 50

4.6 Coverage, quality and recency of evidence syntheses 52

4.7 Living evidence products 55

4.8 Best evidence versus other things (and how to get the most from other things) 57

4.9 Contexts that shape how evidence is viewed 59

4.10 Indigenous rights and ways of knowing 60

4.11 Misinformation and infodemics 62

4.12 Weaknesses in a health-research system 65

4.13 Weaknesses in many COVID-19 evidence-support systems 67

4.14 Features of an ideal national evidence infrastructure 69

4.15 Global-commission reports by form of evidence 72

4.16 Annex to section 4.5 – Examples of quality-assessment tools 73

4.17 References 75

This chapter is the third of three chapters exploring the issue at the heart of this report: what is involved in systematizing the use of evidence, by the full range of decision-makers, in addressing societal challenges? Here we focus on studies, syntheses and guidelines, or the supply of evidence. Chapter 2 focuses on the nature of societal challenges. Chapter 3 focuses on decisions and decision-makers, or the demand for evidence.

Copyright © 2022 McMaster University. All rights reserved. This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License. No part of this report may be adapted in any way without the prior written permission of the publisher.

McMaster University is making this report chapter freely available to advance the public interest. However, the University does not give permission for this chapter to be posted on websites other than the website of the Evidence Commission secretariat (the McMaster Health Forum), or to be circulated electronically. The full report is also available for free downloading from the Evidence Commission website, and for purchase on Amazon and other online stores.

This report chapter and the information contained herein are for informative, public interest purposes only, are provided on an as-is basis (without warranty, express or implied), and are not meant to substitute for financial, legal or medical advice. McMaster University, the Evidence Commission secretariat, the commissioners and the publisher assume no responsibility or liability for loss or damage caused or allegedly caused, directly or indirectly, by the use of information contained in this report chapter, and they specifically disclaim any liability incurred from its use.

Acting on behalf of the Evidence Commission, the McMaster Health Forum welcomes feedback about the report, as well as suggestions about pathways to influence for the report's recommendations. Please send your comments to evidencecommission@mcmaster.ca.

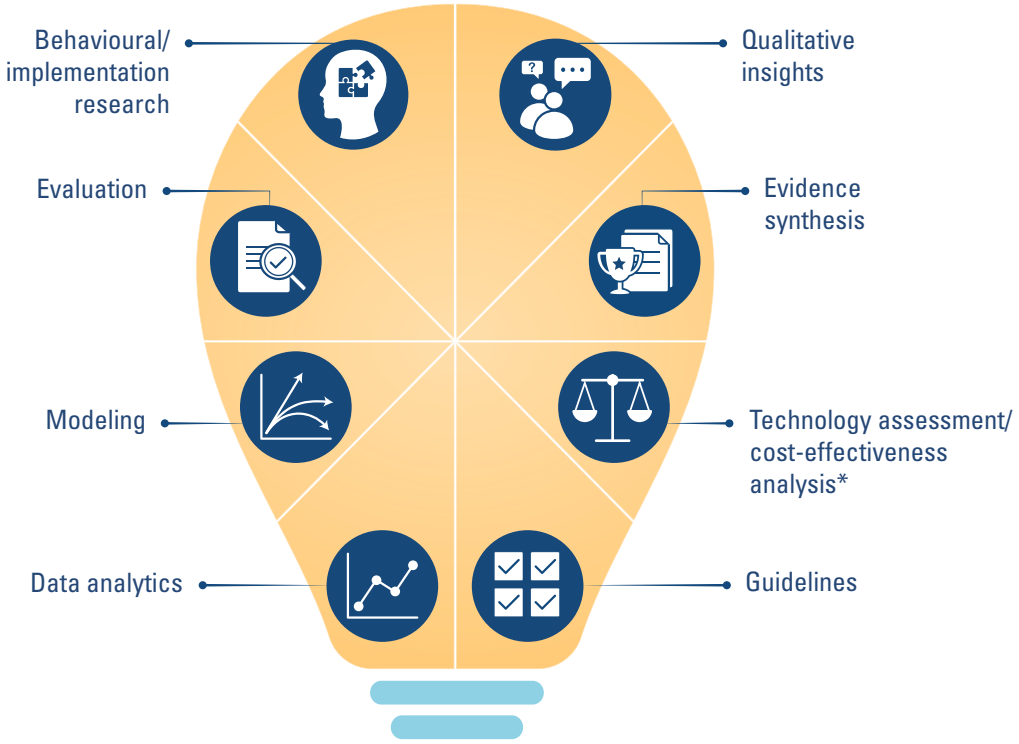
The appropriate citation for this report is: Global Commission on Evidence to Address Societal Challenges. Chapter 4. Studies, syntheses and guidelines: Supply of evidence. The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers. Hamilton: McMaster Health Forum, 2022; p. 43-75.

ISBN 978-1-927565-31-5 (Online)
ISBN 978-1-927565-30-8 (Print)

Global Commission on Evidence
to Address Societal Challenges

4.1 Forms in which evidence is typically encountered in decision-making

Evidence is typically encountered in decision-making in eight different forms. These forms can be interrelated. For example, an evaluation featuring a randomized-controlled trial may also incorporate evidence that draws on data analytics, qualitative insights, and a cost-effectiveness analysis. Similarly, a case study may draw on both qualitative insights about experiences and preferences and quantitative evidence from data analytics, modeling and evaluations.



**We have grouped technology assessment and cost-effectiveness analysis because they are often conducted for the same types of products and services and by the same evidence groups, and because a cost-effectiveness analysis is almost always a key element of a technology assessment. We recognize that the producers of some of these forms of evidence place more emphasis on the process than the resulting evidence product, but these forms of evidence can still be encountered by many decision-makers who have not been involved in any related process.*

The 'studies' referred to in this chapter's title (e.g., an evaluation, a behavioural-research study, a qualitative study, and other forms of 'primary' research) can generate many of these forms of evidence. The 'syntheses' from the chapter title are a form of evidence in their own right and are sometimes called 'secondary' research. The guidelines from the chapter title are also a form of evidence, and as we discuss in **section 4.4**, technology assessments can also include recommendations.

We use the term 'evidence' as a short form for 'research evidence,' recognizing that there are many other types of evidence (e.g., evidence that individuals themselves derive from their own lived experiences and evidence considered in a court of law) and that evidence is one of many factors that can influence a decision. We define each of these terms in **section 4.2** and show how each form of evidence relates to steps in a decision-making process. We describe the reverse – how each step in a decision-making process relates to forms of evidence – in **section 4.3**.

4.2 Definitions of forms in which evidence is typically encountered

We provide below simple definitions of each form of evidence. We have adapted many of these from others' definitions, with the goal of more clearly differentiating the eight forms of evidence while also showing how they interconnect. We also note how each form of evidence relates to any of the four steps in a decision-making process.

Together with **section 4.3**, which describes how each step in a decision-making process relates to forms of evidence, this section builds on the list of decision-making questions first introduced in **section 3.1**.

Forms of evidence	Definitions	Steps where it adds the greatest value			
Data analytics	Systematic analysis of raw data to make conclusions about that information	1			4
Modeling	Use of mathematical equations to simulate real-world scenarios (i.e., what is likely to happen if we don't intervene) and options (i.e., what happens if we intervene) in a virtual environment	1	2		
Evaluation	Systematic assessment of the implementation (monitoring) and impacts (evaluation) of an initiative for the purposes of learning or decision-making				4
Behavioural/ implementation research	Study of methods to promote the systematic uptake of effective approaches into routine practices at the personal, professional, organization and government levels (implementation research) Systematic examination of what people (citizens and professionals) do, what drives them to do it, and what can sustain or change what they do (behavioural research)			3	

Qualitative insights	Study of (typically non-numerical) data – obtained from interviews, focus groups, open-ended questionnaires, first-hand observation, participant-observation, recordings made in natural settings, documents, and artifacts – to understand how individuals and groups view and experience problems, options, implementation considerations (barriers, facilitators and strategies), and metrics	1	2	3	4
Evidence synthesis	Systematic process of identifying, selecting, appraising and synthesizing the findings from all studies that have addressed the same question in order to arrive at an overall understanding of what is known, including how this may vary by groups (e.g., racialized communities) and contexts (e.g., low socio-economic neighbourhoods)	1	2*	3	4
Technology assessment/ cost-effectiveness analysis	Assessment of all relevant aspects of a 'technology' (e.g., a product or service), including safety, effectiveness, and economic, social and ethical implications (technology assessment), with an evidence synthesis often contributing to the assessment of effectiveness Comparison of the relative outcomes (effectiveness) and costs of two or more options, again with an evidence synthesis often contributing to the assessment of effectiveness		2*	3	4
Guidelines	Systematically developed statements that recommend a particular course of action, often for citizens and professionals and sometimes for organizations and governments, with one or more evidence syntheses contributing to the assessment of effectiveness, values and preferences, and other factors		2		

**Adds the greatest value in this step but can add value in other steps*

Note that briefs, infographics, plain-language summaries and other documents derived from any form of evidence or any combination of forms of evidence can be used to package key information for a distinct type of decision-maker. Such 'derivative evidence products' can be used in dissemination and implementation initiatives targeting such decision-makers and add value in all steps.

4.3 Matching decision-related questions to forms of evidence

Having mapped forms of evidence to steps in a decision-making process in **section 4.2**, here we map each step in a decision-making process to forms of evidence, with examples.

Evidence syntheses can help answer almost all of these questions by summarizing what we know and don't know based on all of the studies that have addressed a similar question. Evidence syntheses are critically important for questions about benefits and harms, both for options and for implementation strategies. We elaborate in **section 4.4** on why evidence syntheses are the best place to start when answering many types of questions.

Steps	Related questions	Examples of helpful forms of evidence
1	Indicators – How big is the problem?	Data analytics
	Comparisons – Is the problem getting worse or is it bigger here than elsewhere?	Data analytics *(e.g., using administrative databases or community surveys)*
	Framing – How do different people describe or experience the problem and its causes?	Qualitative studies *(e.g., using interviews and focus groups)*
2	Benefits – What good might come of it?	Evaluations *(e.g., effectiveness studies like randomized-controlled trials)*
	Harms – What could go wrong?	Evaluations *(e.g., observational studies)*
	Cost-effectiveness – Does one option achieve more for the same investment?	Technology assessments / cost-effectiveness evaluations
	Adaptations – Can we adapt something that worked elsewhere while still getting the benefits?	Evaluations *(e.g., process evaluations that examine how and why an option worked)*
	Stakeholders' views and experiences – Which groups support which option?	Qualitative studies *(e.g., using interviews and focus groups to understand what is important to citizens)*
3	Barriers and facilitators – What (and who) will get in the way or help us in reaching and achieving desired impacts among the right people?	Qualitative studies *(e.g., using interviews and focus groups to understand barriers and facilitators)*
	Benefits, harms, cost-effectiveness, etc. of implementation strategies – What strategies should we use to reach and achieve desired impacts among the right people?	Behavioural / implementation research *See also 'selecting an option'*
4	Is the chosen option reaching those who can benefit from it?	Data analytics
	Is the chosen option achieving desired impacts at sufficient scale?	Evaluations

4.4 Interplay of local and global evidence

Decision-makers need both local evidence (i.e., what has been learned in their own country, state/province or city) and global evidence (i.e., what has been learned around the world, including how it varies by groups and contexts). By 'local' we mean national and sub-national, and that evidence can take many forms, including local data analytics, a local evaluation, and local implementation research. The global evidence typically takes the form of an evidence synthesis, which we return to below.

Decision-makers may benefit from recommendations that draw on both local and global evidence. Guidelines, by definition, provide recommendations. As we note in the introduction, in times of crisis we must often initially rely on emerging guidance (e.g., we don't yet know enough but wash your hands well in the meantime) and then on replacement guidance (e.g., we now have evidence indicating that masks reduce transmission). At all times, we need to be open to what have been called 'reversals,' which is when accumulating evidence shows that approaches thought to have benefits turn out to not actually work, or even cause harm. Technology assessments may provide recommendations, or they may provide a type of evidence support by complementing the available evidence with an assessment of the social, ethical and legal factors that may also influence a local decision.

Modeling is most commonly a form of local evidence. However, it can provide a way of synthesizing the best evidence globally, as is done in high-stakes domains like climate action, medicines reimbursement, and macroeconomic policy. Modeling can also provide a form of local evidence support, with modelers effectively acting as a type of evidence intermediary. This was the case with many jurisdiction-specific COVID-19 models that government policymakers drew on to predict the likely future impacts (and most consequential uncertainties) of options like lockdowns. When done well, this modeling used effect estimates from evidence syntheses or, in their absence, systematically elicited expert opinion.

Local and global evidence may be informed or complemented by other forms of analysis, such as policy, systems and political analysis. We discuss these types of analysis in **section 5.4**.

Vantage point	Forms of evidence
Local (national or sub-national) evidence	Data analytics; Modeling; Evaluation; Behavioural/ implementation research; Qualitative insights
Global evidence	Evidence synthesis
Local (national or sub-national) recommendations or evidence support informed by local and global evidence	Technology assessments; Guidelines

Global evidence

An evidence synthesis uses a systematic and transparent process to identify, select, appraise and synthesize the findings from all studies that have addressed the same question. The objective is to come to an overall understanding of what is known, including how this may vary by groups (e.g., girls and young women) and contexts (e.g., low- and middle-income countries). For questions about options, part of what is known can be about what works for whom in what contexts.

An evidence synthesis offers four advantages over other approaches to summarizing the best evidence globally, such as an expert conducting an informal narrative review of the scientific literature:

- Reduces the likelihood of being misled by ensuring that all relevant studies have been included and that greater weight is given to high-quality studies
- Increases confidence about what can be expected by increasing the number of study participants included in the analysis
- Makes it easier to assess what the global evidence means in a particular context by presenting information about the participants and contexts being studied, and ideally how the findings varied according to such factors
- Makes it easier to contest the available evidence by ensuring that everyone has access to the same 'data' and clear reporting about how the data were synthesized.

The first of these advantages can help to address what is sometimes called the replication or reproducibility crisis in science – many findings from a single study cannot be replicated or reproduced. The crisis has been documented in many fields from medicine (e.g., hydroxychloroquine and ivermectin to treat COVID-19) to economics and psychology. More troubling is the fact that non-replicable findings are cited more than replicable ones, even after the failure to replicate has been documented.(1)

Model-based explorations of the future to address the 'complexity cubed' societal problem of climate change, using multiple types of evidence and drawing on robust intercomparison exercises, provides an alternative paradigm to the type of evidence synthesis described above. Building on the best of both approaches could be a fruitful way forward.(2)

Local evidence

Local (national or sub-national) evidence can shed light on whether there's a local problem and its causes, on the local feasibility and acceptability of an option to address a problem, and on local factors that may get in the way or help in reaching and achieving desired impacts among the right people. What 'local' means for decision-makers will vary – for one person 'local' may be their country; for another, it may be their immediate neighbourhood. We address the issue of the local applicability of evidence in **section 4.5**.

4.5 Distinguishing high- from low-quality evidence

Not all evidence is high quality and reliable for making decisions. Tools exist for many (but not all) forms of evidence to help make judgements about whether the evidence (from a single study or a body of evidence) can be relied upon. As we describe here, these tools use scores or grades to help users understand how confident they can be in the evidence. Many journals now require authors to follow reporting standards, such as CONSORT for randomized-controlled trials and PRISMA for evidence syntheses. Most journals do not require reviewers to use specific tools to assess the quality of studies or strength of recommendations; as a result, publication in a peer-reviewed journal is not a good proxy for quality.

Issue	Response
Studies (and guidelines) vary in their quality (or trustworthiness)	• Quality-assessment (or critical-appraisal) tools have been developed for specific study designs (e.g., randomized-controlled trial), for broad categories of study designs (e.g., observational study, qualitative research, and evidence synthesis), and for guidelines – see the annex at the end of this chapter (**section 4.16**) for examples (RoB2, ROBINS-I, JBI checklist, AMSTAR, and AGREE II) • Tools may yield a summary judgement (e.g., low risk of bias using RoB2 or ROBINS-I), a score that some group into ranges (e.g., high quality using AMSTAR), a set of scores (e.g., six domains using AGREE II), or a set of considerations that can inform a summary judgement (e.g., JBI checklist)
Bodies of evidence vary in their certainty (or the confidence you can place in them)	• Certainty-assessment tools have been developed for a body of evidence addressing the same question (e.g., effect of an intervention on a specific outcome or the meaning that citizens attach to a particular phenomenon) – see **section 4.16** for two examples (GRADE and GRADE CERQual) • Tools may yield a summary judgement about confidence that the true effect is similar to the estimated effect (e.g., high certainty with GRADE) or that the phenomenon of interest is well represented by a qualitative study finding (with GRADE CERQual) • A summary judgement about the certainty of an effect estimate is more helpful than a test of statistical significance demonstrating that an intervention 'works' or 'doesn't work' (which will happen by chance one in 20 times if statistical significance is set at the 0.05 level)
Recommendations vary in their strength	• Strength-assessment tools have been developed for guideline recommendations (e.g., GRADE, in addition to ranking the certainty of a body of evidence, as described above) – see **section 4.16** for an example • Tools may yield a summary judgement about whether most decision-makers would choose to proceed with an intervention (e.g., strong with GRADE) or whether most would need to carefully weigh the pros and cons of an intervention

Some sources of (or approaches used to generate) evidence can be hard to judge	• No widely accepted tools exist to assess how much confidence can be placed in:
	  ◦ An expert, although examples like The Good Judgement Project do exist for forecasting (we return to expert opinion later in this chapter and, in the case of expert opinion about model parameters, in **section 4.16**)
	  ◦ Models used in generating some forms of evidence (which we address in **section 4.7** when talking about climate-change models and in **section 4.16**)
	  ◦ An artificial-intelligence algorithm used in generating some types of evidence, although examples like TRIPOD are starting to emerge (3)

Distinguishing high- from low-quality evidence is particularly challenging when evidence is embedded in dashboards, models and other formats, and when conflicts of interest are at play. We return to the latter in **sections 4.12**, **4.14** and **4.16**. While not the focus of this report, distinguishing high- from low-quality 'raw data' can also be challenging, and organizations like UNICEF have developed data-quality frameworks to assist with this (bit.ly/3DQQRRv).

Some 'one-stop shops,' such as Social Systems Evidence and the COVID-19 Evidence Network to support Decision-making (COVID-END) inventory (described in **section 4.6**), use some of these tools so that decision-makers and those supporting them can focus on high-quality evidence syntheses or understand that they are using the best available (if not high-quality) evidence syntheses.

The COVID-19 pandemic required decision-makers to make difficult decisions in short time frames, initially with little and often indirect evidence, and then, over time, with studies, bodies of evidence, and recommendations developed using a robust process. To support decision-making about COVID-19 based on bodies of evidence (rather than single studies), COVID-END profiled in its inventory of 'best' evidence syntheses those that were up-to-date (based on the date of searching for evidence), were high quality (based on the AMSTAR tool), and provided an assessment of the certainty of the evidence (based on the GRADE tool).

Just as not all evidence is high quality, not all global evidence will be applicable in a given context. For example, an evidence synthesis containing studies conducted in only high-income countries may have limited applicability to some low-income countries. There may be important differences in baseline conditions, in on-the-ground realities and constraints, and in structural features of the local system (e.g., national health system or provincial/state education system). A SUPPORT tool can also help people think through the local applicability of findings from an evidence synthesis and consider how insights can still sometimes be drawn even when the findings aren't applicable.(4)

Bayesian reasoning has garnered increasing attention as a way to deliberately re-draw our 'mental maps' about challenges and ways of addressing them, not by replacing all of what we thought we knew with new information, but by modifying our understanding to an appropriate degree. The degree depends on how much confidence you had in your pre-existing knowledge (the 'prior' probability of something being true) and how much confidence you place in the new knowledge. More confidence can be placed in the new knowledge if it comes from a high-quality evidence synthesis that includes studies conducted in contexts similar to your own.

Evidence intermediary and producer, Gillian Leng

Experienced executive leading a technology-assessment and guideline agency that supports health and social care decision-making by governments, service providers and patients

The UK has led work over many years to encourage the synthesis and use of evidence – from the first randomized-controlled trial to prevent scurvy in sailors, to the more recent innovative What Works Centres to promote the use of evidence in a range of policy areas. As part of this evidence-based movement, over the last 20 years the National Institute for Health and Care Excellence (NICE) has transformed the use of evidence in healthcare practice, as well as in wider public-health initiatives and social care.

The COVID-19 pandemic has dramatically reinforced the need for high-quality evidence to inform policy and practice, and has also highlighted the negative consequences of social media and associated misinformation. In this context, the work of the Global Commission on Evidence to Address Societal Challenges is hugely important, and should be seen as essential reading for all policymakers around the world.

4.6 Coverage, quality and recency of evidence syntheses

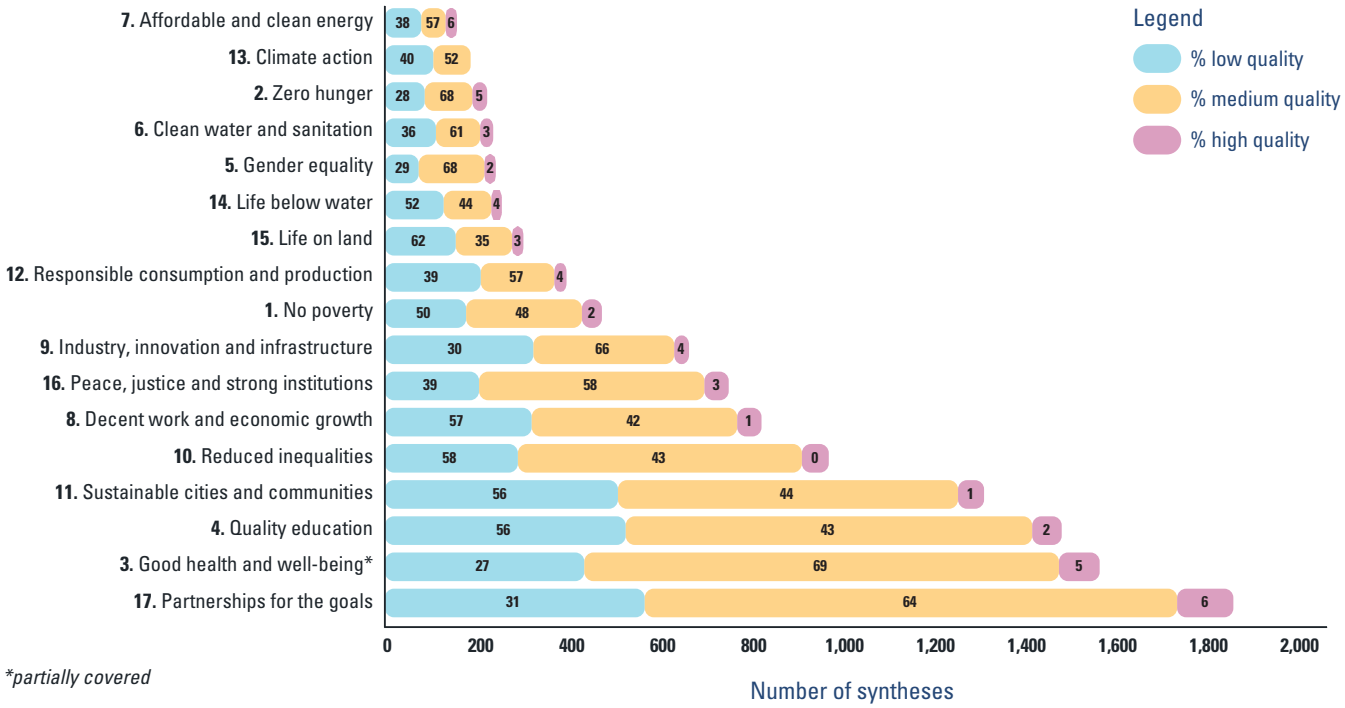
The global stock of evidence syntheses suffers from incomplete coverage of priority topics, a wide spectrum of quality (of the synthesis), and problems with recency (of the search for potential studies to be included in the synthesis). Analyses of two 'one-stop shops' for evidence syntheses illustrate the magnitude of the problem. One 'shop' focuses on all of the non-health Sustainable Development Goals, or SDGs (Social Systems Evidence), and the other focuses on all potential COVID-19 responses (COVID-END inventory of best evidence syntheses and the larger database from which the inventory is drawn).

SDG evidence syntheses

Of the 4,131 SDG evidence syntheses – defined as overviews of reviews, reviews of effects, and reviews addressing other questions – included in Social Systems Evidence as of 12 August 2021:

- coverage was uneven, with seven SDGs addressed by a relatively small number of evidence syntheses (263 or fewer) relative to the number of questions that can be asked in relation to each SDG (2 – Zero hunger, 5 – Gender inequality, 6 – Clean water and sanitation, 7 – Affordable and clean energy, 13 – Climate action, 14 – Life below water, and 15 – Life on land)
- quality was uneven, with seven SDGs addressed by a stock of evidence syntheses in which at least half are of low quality (6 – Clean water and sanitation, 7 – Affordable and clean energy, 9 – Industry, innovation and infrastructure, 12 – Responsible consumption and production, 13 – Climate action, 14 – Life below water, and 15 – Life on land)
- all SDGs have a median year of last search that is five or six years ago (2016 or 2017)
- only between one in 10 (12%) and one in five (21%) evidence syntheses about most SDGs included at least one study from a low- and middle-income country, with an even lower percentage (3%) for one SDG (9 – Industry, innovation and infrastructure).

The number and quality of evidence syntheses are presented by SDG in the bar chart below.



partially covered

Keep the following in mind with this bar chart:

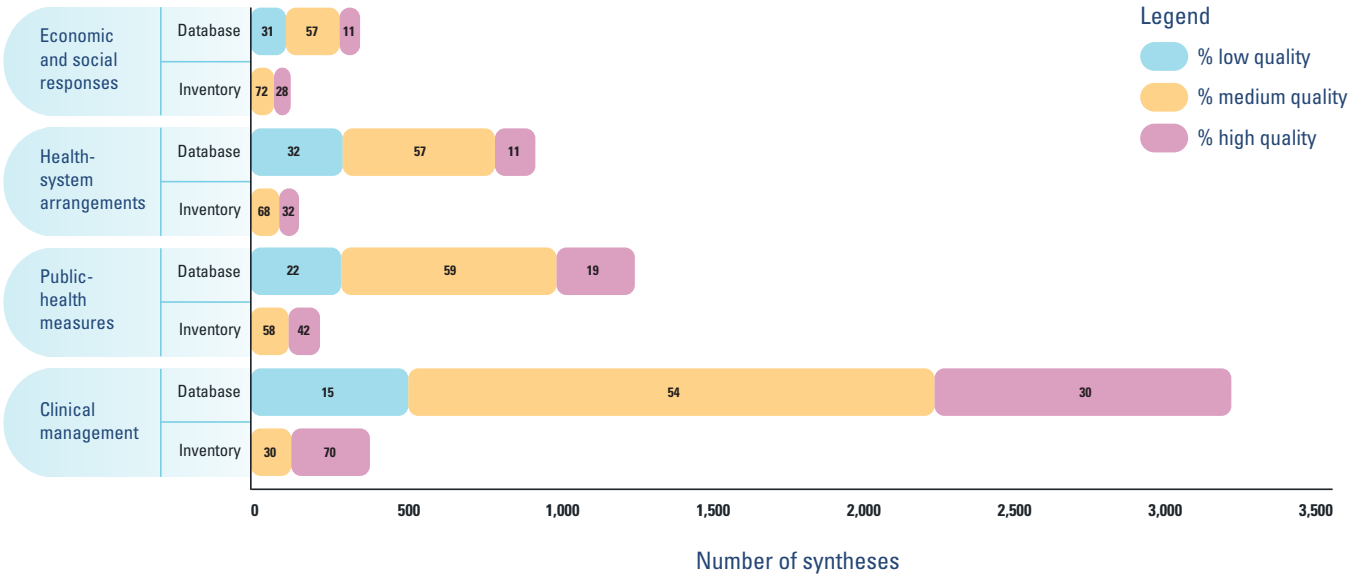
- the numbers add to more than the total number of evidence syntheses because a synthesis may address more than one SDG
- the number of evidence syntheses addressing:
  - SDG3 is a significant undercount, with health-related evidence syntheses included only if they also address another SDG
  - SDG17 is a significant overcount, with many evidence syntheses addressing another SDG as their primary question also addressing partnerships as a secondary question
  - SDGs 7, 13, 14 and 15 may be an undercount as they have been a more recent focus for inclusion in Social Systems Evidence
- quality ratings have been completed for 85% of the evidence syntheses included in Social Systems Evidence.

COVID-19 evidence syntheses

Of the 4,256 COVID-19-related evidence syntheses included in the full COVID-19 database and the 562 COVID-END inventory of best evidence syntheses, as of 1 August 2021:

- coverage was uneven, with only 237 evidence syntheses addressing economic and social responses to COVID-19 (of which only 49 were included in the inventory), while much higher numbers addressed clinical management (3,128), public-health measures (1,148), and health-system arrangements (818)
- quality was uneven, with roughly one quarter (26%) of COVID-19 evidence syntheses being low quality and over half (56%) being medium quality
- three of the four COVID-19 response categories have a median date of last search that is within 4.5 months of the World Health Organization (WHO) declaring a pandemic (11 March 2020).

The much more recent median search date for clinical management – 12 months after the pandemic declaration and 4.5 months before the analysis was completed – was driven by the large number of comparisons of drug treatments, all with the same search date, on the COVID-NMA living evidence platform. The number and quality of evidence syntheses are presented by broad category of COVID-19 response in the bar chart below.



Keep in mind the following with this bar chart:

- the numbers add to more than the total number of evidence syntheses because a synthesis may address more than one category of the COVID-END taxonomy
- evidence syntheses needed to have a quality rating of medium or high to be considered for inclusion in the COVID-END inventory of 'best evidence syntheses.'

These findings echo similar shortfalls in the stock of evaluations (specifically randomized-controlled trials), evidence syntheses, and evidence maps (of evaluations and evidence syntheses) available to inform decision-making about:

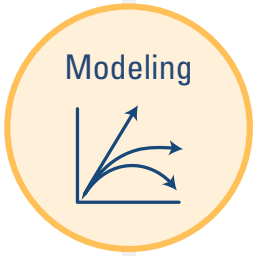
- education, where only 25% of trials had more than 1,000 participants (and only 12% of trials conducted in the 1980-2016 period were performed in Asia, Africa or Central and South America) (5)
- health, where only 16% of evidence syntheses incorporated quality assessment in their analysis (although 70% conducted such as assessment) and more generally reporting quality was highly variable (6)
- sustainable development in low- and middle-income countries, where four or fewer evidence maps reported outcomes relevant to eight of the 17 SDGs in the 2010-17 period, and one quarter of the evidence maps did not address equity in any way.(7)

Other such stock-taking exercises have been framed more positively, such as the one noting that the 740 randomized-controlled trials in social work demonstrate that this approach to evaluation is indeed possible in the field.(8)

4.7 Living evidence products

Four of the forms of evidence that decision-makers typically encounter are now available as 'living' evidence products, meaning they are regularly updated as new data are added or new studies are published. Many such living evidence products began as part of the COVID-19 evidence response. Fewer exist in sectors other than health. We provide examples below.

Many government policymakers and other decision-makers have come to expect such regular updating for COVID-19 and will likely start to ask why such products can't be maintained for other high-priority societal challenges where there is significant uncertainty and a high likelihood of evidence emerging to address that uncertainty. The growing use of artificial intelligence, among other innovations, will likely make it easier in the future for evidence producers to meet these greater expectations. However, evidence producers will need to take steps to ensure that these innovations do not inadvertently perpetuate or increase the risk of discrimination (e.g., using race or variables associated with race in ways that disadvantage certain groups). They will also need to support decision-makers to interpret and use the findings appropriately, especially when causal inferences are being made.

Forms of evidence	Examples of living evidence products
	• The WHO COVID-19 Dashboard provides a set of data analytics about the stringency of public-health measures being taken to address COVID-19, the UK Health Security Agency surveillance reports (bit.ly/3DeaSIc) provide a set of data analytics about COVID-19 in the UK, and Opportunity Insights' Economic Tracker provides a set of data analytics about COVID-19 impacts on the economic prospects of people, businesses and communities in the US • The Organisation for Economic Co-operation and Development (OECD) Weekly Tracker of Economic Activity provides a set of data analytics about economic activity for most OECD and G20 countries
	• The European COVID-19 Forecast Hub presents every week a forecast of cases and deaths per week per 100,000 people – both overall and by country – based on an ensemble of models, while the Institute for Health Metrics and Evaluation COVID-19 Projections updates every two weeks a model of projected deaths from COVID-19, both those reported as COVID-19 and those attributed to COVID-19, that could be used to explore a range of scenarios (e.g., about mask use and vaccine uptake) in specific countries • The Intergovernmental Panel on Climate Change presents every five-to-seven years an assessment report that draws on modeling of human-induced climate change, its impacts, and possible response options, although strictly speaking this is a synthesis of findings from models (which may or may not be living) informed by a robust process of inter-model comparisons (which is undertaken by different scientists for each assessment report – see bit.ly/3wKQy8D for an example)
	• COVID-END living evidence synthesis #6 provides updates every two weeks about COVID-19 vaccine effectiveness against variants, and COVID-NMA updates weekly evidence syntheses about all drug treatments for COVID-19 (and later added preventive therapies and vaccines) • The Global Carbon Project updates annually, based on modeling and empirical studies, estimates of the five major components of the global carbon budget (anthropogenic carbon-dioxide emissions and their redistribution among the atmosphere, ocean and terrestrial biosphere in a changing climate) and their associated uncertainties
	• The Living WHO Guideline on Drugs for COVID-19 provides updates every one-to-four months about COVID-19 drug treatments, and the National COVID-19 Clinical Evidence Task Force updates weekly evidence-based COVID-19 guidelines for Australian health professionals • The Education Endowment Foundation maintains living guidance for schools as part of their Teaching and Learning Toolkit, such as the one addressing teaching-assistant interventions

A thematic analysis of a listserv discussion among the COVID-END Community identified differing views about:

- What is understood by the term 'living' evidence synthesis (e.g., can the spectrum of 'living' status be better captured using a scale than a yes/no designation, and should a minimum threshold be set for frequency of updates)
- When one should be started or when an existing synthesis should become 'living' (e.g., new evidence is rapidly becoming available, and that evidence is likely to address key areas of uncertainty among decision-makers about a topic of high priority to them)
- When updates can be stopped (e.g., evidence is unlikely to change interpretations about what we know, and the priority accorded to the topic is downgraded)
- Where and how one can best be disseminated (e.g., can journals accommodate a process where an initially peer-reviewed synthesis is updated regularly without the delay of additional peer review, and can decision-makers rely on commitments to provide updates at defined times)

Such issues will likely be the focus of intense debate in the coming years. Additional details about the rationale for living evidence syntheses and the issues involved in maintaining them can be found in a brief note co-authored by one of our commissioners.(9)

In **section 4.13**, we describe some of the key characteristics of the living evidence syntheses maintained as part of the COVID-19 evidence response.

Evidence producer, Jan Minx

Impact-oriented scholar bringing innovative evidence-synthesis approaches to domestic policy advice and global scientific assessments about climate change and sustainability

I am working at the interface between two forms of evidence: 1) evidence syntheses, which seek to learn from the past and are widely used in the health sector; and 2) modeling, which seeks to predict the future and is widely used in the field of climate change. I strongly support **recommendation 19** – we need to learn from evidence groups in other sectors. As we note in that recommendation, Cochrane has pioneered many approaches to synthesizing studies about what works in health, including living evidence syntheses, and the Intergovernmental Panel on Climate Change (IPCC) has pioneered many approaches to modeling human-induced climate change over long time horizons. Cochrane and the IPCC can learn from each other and from others, and others can learn from them.

4.8 Best evidence versus other things (and how to get the most from other things)

Many individuals and groups bring forward what they call evidence to address societal challenges. 'Best evidence' in a given national (or sub-national) context – in the form of national (or sub-national) evidence drawn from the best available studies (i.e., what has been learned in that context) and global evidence drawn from the best available evidence syntheses (i.e., what has been learned from around the world, including how it varies by groups and contexts) – needs to be differentiated from 'other things' that are sometimes presented as evidence, such as a single study, expert opinion, an expert panel, a research interest group, an anecdote 'dressed up as a case study,' a white paper, and a jurisdictional scan. Each of these other things brings with them a risk (column 2 below). At the same time, there are ways to get more value from them (columns 3 and 4 below).

We do not consider here 'other things' beyond those typically presented as research evidence, such as people's lived experiences (which we discuss in **section 2.3** in the context of co-designed interventions) or Indigenous ways of knowing (which we discuss in **section 4.10** as part of a broader discussion about Indigenous peoples).

If presented with…	…which brings with it a risk of…	…then…	…or better yet…
Single study *(including preprints)*	'Hubcap chasing,'* or giving attention to each study that is actively promoted by the authors, their media-relations office or others (as happened with the high-risk-of-bias study about hydroxychloroquine discussed in **section 3.7** and the now retracted study** about a link between vaccines and autism)	Ask for a critical appraisal of the study using widely accepted quality criteria (to understand the risk of bias) and recognize that a statistically significant finding (at the 0.05 level) may be found by chance in one in 20 studies	Add the study to a 'living' evidence synthesis where it can be understood alongside other studies addressing the same question (or consider it as one of many types of national or sub-national evidence to be put alongside the best global evidence)
Expert opinion	'Squeaky wheel getting the grease' / 'eminence-based' (rather than evidence-informed) decision-making, or giving attention to those who command the greatest attention by virtue of persistence, reputation or other factors (as happened with widely viewed television shows about the Scared Straight crime-prevention program even after evidence syntheses*** had found evidence of harm and no evidence of benefit)	Ask the expert to share the evidence (ideally evidence syntheses) on which the opinion is based, as well as the methods used to identify, assess, select and synthesize it	Engage the expert in working through what specific evidence syntheses mean for a specific jurisdiction, or in challenging ways of thinking with different forms of evidence**** (or ask the expert what evidence would convince them they were wrong)
Expert panel	GOBSATT, or 'good old boys sitting around the table' offering their personal opinion	Ask the panel members to share the evidence (ideally evidence syntheses) on which their input and recommendations are based, as well as the methods used to identify, assess, select and synthesize it	Add methods experts to the panel (or secretariat), pre-circulate the best local (national or sub-national) and global evidence, support robust deliberation, and make explicit which recommendations are based on what strength of evidence

* *We use the term 'hubcap chasing' (i.e., dogs repeatedly barking at and chasing cars) as a metaphor for sharing and commenting upon each new study that captures one's attention.*

** *www.nature.com/articles/nm0310-248b*

*** *onlinelibrary.wiley.com/doi/10.4073/csr.2013.5*

**** *Such challenges have been called 'red teaming' in the military.*

Jurisdictional scan	'Groupthink,' or people in many jurisdictions relying on people in one jurisdiction who are willing to share their experiences and innovations, but haven't yet evaluated them	Ask or look for any available supporting evidence or plans for generating it	
Research interest group*****	Researchers advocating for action based on their personal values and preferences or their professional interests	Ask groups why their values and preferences should count more than the citizens we all serve	Encourage them to base their requests on high-quality evidence syntheses
'Case study'	Anecdotal experiences given a name that implies a rigorous approach underpins it	Ask the writer what criteria were used to select the case, what mix of data-collection approaches were used, and what analytic and other approaches were used to ensure rigour	
White paper	Taking at face value the implicit or explicit assertion that evidence was used in arriving at a statement of policy preferences	Ask government leaders or advisors to share the evidence they used as a basis for their input and recommendations, as well as the methods used to identify, assess, select and synthesize it	

***** *Note that societal interest groups may also invoke evidence in advocating for action based on their values and preferences, in which case the same response as in column 4 may be appropriate.*

4.9 Contexts that shape how evidence is viewed

Historical, social and cultural contexts can shape how evidence is viewed by, for example, racialized communities (the R in PROGRESS-Plus, which we introduced in **section 1.7**) and by women (the G in PROGRESS-Plus), among others. Some contexts relate directly to past efforts to generate evidence, while others relate to past efforts to portray specific groups as 'different,' which may then manifest as these groups being skeptical about any evidence purporting to be for or about them. These contexts need to be understood if we are going to produce and communicate evidence in ways that will be acted upon.(10; 11)

As we return to in **section 4.10**, contexts, as well as their distinct rights and ways of knowing, can also shape how evidence is viewed by Indigenous peoples. Context can also shape whether and how misinformation flourishes, which is the focus of **section 4.11**.

Examples of contexts	Potential implications for how evidence is produced and communicated
Directly related to past efforts to generate evidence in the US	
Effective treatment was withheld from Black men with syphilis so that the progression of untreated syphilis could be monitored (bit.ly/3DeaH9x)	Give greater attention to what is (and is not) examined, by whom it is examined (e.g., research teams comprised of people drawn from different contexts), how it is examined (e.g., more participatory approaches that are ethically grounded and equity oriented), and why it is examined (e.g., to identify strengths to be built upon)
Trials of treatment for heart disease did not include women yet the findings were assumed to apply to them (bit.ly/3olxgTH)	
Standardized testing of students has been done in ways that disadvantaged students of colour, particularly those from low-income families (bit.ly/3wDICGk)	
Related to past efforts to portray specific groups as 'different' in their newly adopted countries	
False depictions of Chinese immigrants as dirty and diseased were used to justify the particularly strict enforcement of sanitary regulations in their San Francisco community (bit.ly/3qzeJFV)	Give greater attention to how evidence is portrayed in various media and draw on these insights in seeking to anticipate how groups will respond to evidence for or about them, or to understand why they are responding in the way they are
Implicit messages about Black people in Thatcher-era Britain being an 'external' source of the country's problems appeared in books and films and were accepted as true by some audiences (bit.ly/3naBa2n)	
Media coverage framed certain populations such as Muslim immigrants to Europe and Iraqi detainees after the US invasion of Iraq as already 'lost' (to unemployment, starvation and prison) and not worthy of societal protection (bit.ly/3wGrKyE)	

4.10 Indigenous rights and ways of knowing

As part of a broader shift to recognize and ensure the rights of Indigenous peoples, many government policymakers, researchers and others are coming to accept that Indigenous people should have control over data-collection processes, and that they should own and control how this evidence is used. Building on the First Nations data principles of ownership, control, access and possession (sometimes called the OCAP principles), the International Indigenous Data Sovereignty Interest Group developed the CARE Principles for Indigenous Data Governance (with CARE capturing the first letters of collective benefit, authority to control, responsibility, and ethics). These principles were designed to complement the FAIR guiding principles for scientific data management and stewardship (with FAIR capturing findable, accessible, interoperable, and reusable). The goal is that stewards and users of Indigenous data will be 'FAIR' and 'CARE.' Such evidence-related rights should be understood as part of a much broader set of rights established through the United Nations Declaration on the Rights of Indigenous Peoples.

Indigenous ways of knowing is a term that reflects the diversity and complexity of Indigenous approaches to learning and teaching. The diversity arises from the many Indigenous peoples or nations that developed their own ways of knowing, ways that evolved over centuries before the colonization of their lands began, and in the time since then. The complexity arises from many factors, including the many sources of knowledge. While there are commonalities among Indigenous forms of knowledge (e.g., a holistic view of individuals as being interconnected with the people around them and with the land), it is best never to generalize. The table here was developed under the guidance of commissioner Daniel Iberê Alves da Silva (of the M'byá Guarani people), whose biography appears in **appendix 8.2**, as an entry point for discussions about Indigenous ways of knowing. Further discussions should always be led by Indigenous people, as was this one.

Domains	Details
Sources of Indigenous ways of knowing	• Knowledge comes from the relationships of the individual with the world, which has both a material dimension and an inseparable spiritual dimension • Sources of knowledge include plants, animals, other humans, and elements of the land (such as mountains and rivers), as well as dreams, spirits and other manifestations of the spiritual world • The world of water, for example, includes lakes and rivers and also the spirits that inhabit them. More generally the physical territory where a culture of Indigenous people was born and developed over centuries is inhabited by many 'things' that possess spirit, making them 'beings' (and this makes forced relocation particularly damaging) • The physical environment can serve as a prompt or inspiration for the spiritual dimension to help shape a course of action (e.g., watching a river flow can allow an approach for addressing an issue to come to the watcher) • Learning comes from doing alongside someone who holds the knowledge about the 'secret' in how to do it
Characteristics of Indigenous ways of knowing	• Indigenous knowledge is holistic and connected with the history, culture and territory of each people (e.g., their creation stories and how they relate to other 'beings') • 'Knowing' manifests itself in the experiences or 'being' of individuals (e.g., rites of passage are processes in which the experience of discovering the nature of things is 'lived' by individuals) • Knowledge is shared within and across Indigenous peoples and with others, and is refined over time (e.g., a canoe is made differently today than it was two centuries ago) • Knowledge can be acquired through the use of one's own senses (in the traditional sense of the physical senses, but also through clothing, diets, drawings and songs) and through both speaking (what can be said) and contemplation (what cannot be said)
How 'things' are classified within Indigenous ways of knowing	• Categories are perceived differently by different individuals and by different Indigenous peoples in relation to their culture, history or territory (e.g., a plant may be classified one way by one Indigenous people based on its use in healing, and by another based on its association with death) • Categories can shift over time (e.g., some plants were once people) and be understood in terms of their intrinsic 'spirit'

How Indigenous ways of knowing are passed on

- Indigenous knowledge can be transmitted orally (spoken words but also chanting, gestures and silence), by a 'way of being' (learning by doing as well as contemplation), and by the 'memory of things' (narrative history)
  - A story keeper may combine the memory of things and chanting to deliver the right chant – from among hundreds – for the right occasion and at the right time
- Knowledge holders safeguard and share the knowledge in a specific territory (e.g., the medicinal value of a local plant) and do so in a way that emphasizes common purpose (over individual gain), charitable purpose (over power or domination), and ethical purpose (over hoarding the knowledge)
- Learning may also come from the 'beings' in the forest (e.g., animals and rivers)

Relationships of Indigenous ways of knowing to Indigenous worldviews

- Each Indigenous people has their own worldview, while Indigenous peoples also share worldviews that bring them together
- Worldviews can be forgotten, erased, denied and borrowed, as well as constructed for the cultural resistance of today's Indigenous peoples
- Worldviews and forms of knowledge are intrinsically intertwined; Indigenous peoples interpret their 'worlds' from their diverse forms of knowing and knowledge

Relationships of Indigenous ways of knowing to scientific ways of knowing

- The knowledge of each people is in its own physical and spiritual territory, and this knowledge has often been taken from Indigenous peoples without acknowledgement
- Scientists need to learn to recognize, coexist with and respect Indigenous knowledge in all its complexity and diversity
- Government policymakers and other decision-makers need to recognize that science is sometimes being misused to advance the violation of Indigenous territories, including with deforestation and other activities that threaten the future of Indigenous peoples

4.11 Misinformation and infodemics

Misinformation is false information that is spread, regardless of intent to mislead. Disinformation is the intentional spreading of misinformation. For example, a political opponent or foreign government may engage in a disinformation campaign to achieve a particular goal, such as an electoral advantage or undermining of trust in democratic institutions, independent media, and scientific knowledge. Organized groups may pursue other goals, such as making money or advancing an ideology. Because intent can be very difficult to prove, we use the term misinformation here. While misinformation has been with us for centuries, the internet has transformed its scale, drivers and consequences, as well as possible responses to it.

During the COVID-19 pandemic, people began to use the term 'infodemic' (or 'mis-infodemic') to capture the parallel between the rapid spread of the virus and the rapid spread of misinformation about both COVID-19 and measures to prevent it, manage it, and mitigate its economic and social impacts. Existing misinformation efforts related to vaccines were often re-directed to COVID-19 vaccines once they became available, and many new anti-vaccine efforts were launched.

In 2020, the Broadband Commission for Sustainable Development – sponsored by the International Telecommunication Union (ITU) and United Nations Educational, Scientific and Cultural Organization (UNESCO) – published a report about countering digital misinformation while respecting freedom of expression.(12)

The report describes five stages in the misinformation life cycle:

- Instigators and beneficiaries, where questions arise about motivation (and goals as described above)
- Agents, where questions arise about techniques, such as bots and fake accounts or false identities
- Messages, where questions arise about formats, with three of the common ones being:
  - emotive claims and narratives, which often mix emotional language, lies or incomplete information, personal opinions, and elements of truth
  - fabricated, de-contextualized or fraudulently altered images and videos, as well as synthetic audio
  - fabricated websites and polluted datasets
- Intermediaries, where questions arise about platforms (e.g., dark web, social media, messaging, and news media) and the platform features that are being exploited (e.g., algorithms and business models)
- Targets and interpreters, where questions arise about who is affected (e.g., individuals such as citizens, scientists and journalists; organizations such as research centres and news agencies; communities such as Black communities and Indigenous peoples; and systems such as electoral processes) and how they react (e.g., ignoring or sharing to debunk the misinformation)

The report distinguishes misinformation from parody and satire, which can both mislead those without the capacity to identify them and counter misinformation by highlighting its absurd elements.

The Broadband Commission for Sustainable Development report also presents potential responses to misinformation and notes examples of intersections with freedom-of-expression rights. The UNESCO report notes the potential complementarity of these responses and the need to ensure the alignment of any responses used.

Monitoring and fact-checking

- Includes monitoring and exposing misinformation (e.g., debunked claims) and fact-checking new claims
- Judgement of trained professionals employed by independent organizations, even when helped by automation, can mitigate the risk of infringing on freedom-of-expression rights

Credibility labeling

- Includes content-verification tools, web-content indicators, signposting (pointing to credible evidence sources), and website-credibility labeling

Educational

- Includes developing citizens' media and information literacy (e.g., critical-thinking and digital-verification skills), as well as journalists' information literacy

Curatorial

- Includes pointing users to official credible evidence sources, and can be used by news media, social media, messaging and search platforms
- Can be misused as a form of private censorship

Technical and algorithmic

- Covers a spectrum from human learning to machine learning and other artificial-intelligence approaches to identify misinformation, provide additional context, and limit spread
- Automation of appeal processes can infringe on freedom-of-expression rights

Counter-misinformation campaigns

- Includes specialized units to develop counter-narratives to challenge misinformation and mobilizing online communities to spread high-quality evidence

Normative

- Includes public condemnations of acts of misinformation and recommendations to address them, often by political and societal leaders

Economic

- Includes advertising bans, demonetizing specific content (e.g., COVID-19 content) and other approaches to remove incentives for misinformation

Legislative and other policy

- Includes criminalizing acts of misinformation, directing internet communication companies to take down content, and providing material support for credible information sources
- Can be misused to weaken legitimate journalism and infringe on freedom-of-expression rights

Investigative

(which can inform legislative and other responses)

- Examines the instigators, degree and means of spread, money involved, and affected communities

The report does not address the evidence underpinning these responses, although many such evidence syntheses exist. For example, one medium-quality, older synthesis (AMSTAR rating 7/11 and search date of 2017) found that correcting misinformation (i.e., response type 1) has a moderate influence on belief in misinformation (with greater effects in health than marketing or politics), rebuttals are more effective than forewarnings, and appeals to coherence are more effective than fact-checking and appeals to credibility.(13) Our aim here is not to provide the current state of knowledge about these responses, or to explore the psychology of misinformation that may underpin them, but to note that evidence syntheses on misinformation responses exist and living evidence syntheses are needed. Such syntheses could provide an evolving understanding of what is known, including how this may vary by groups (e.g., among those who are more susceptible to misinformation or hold particular belief systems) and contexts (e.g., polarized societies).

As we discussed in the introduction, if we can continue building the capacity, opportunity and motivation to use evidence (in this case to address misinformation about societal challenges), while also exercising judgement, humility and empathy, the combination will serve us well. Even when we can rely on both the rigorous testing and reliable self-correcting systems that typically operate in the health sector, we can do better. As Ross Douthat observes in his memoir about living with Lyme disease, we need more people and institutions with a worldview that both: 1) "accepts the core achievements of modern science, treats populist information sources at least as skeptically as it treats establishment sources and refuses to drink the … Kool-Aid"; and 2) "recognizes that our establishment fails in all kinds of ways, that there's a wider range of experiences that fits within the current academic-bureaucratic lines…."(14) Most of us have benefited tremendously from fields like medicine that combine rigorous testing and fairly reliable self-correcting systems. But some – like Ross Douthat – have not. He notes that, "I am more open-minded about the universe than I was seven years ago, and much more skeptical about anything that claims the mantle of consensus. But I am trying not to let that mix of open-mindedness and skepticism decay into a paranoid-outsider form of groupthink."(14)

4.12 Weaknesses in a health-research system

Prior to the start of the COVID-19 pandemic, a group of researchers documented the weaknesses in the health-research system. They called for a reorganization of the system, including the structures (e.g., global collaborations like Cochrane) and incentives (e.g., from universities, funders and journals) that underpin it, in order to better meet the needs of decision-makers.(15-17) They were primarily concerned with three of the forms of evidence that decision-makers typically encounter, namely primary research (and specifically evaluation, especially randomized-controlled trials), evidence syntheses, and guidelines (and to a lesser extent technology assessments).

While some of the weaknesses became more apparent through the COVID-19 evidence response, the pandemic response also generated notable examples of efforts to address many of the weaknesses. Although the researchers were originally focused on health challenges and on select forms of evidence, many of the insights also apply to other societal challenges and to other forms of evidence. That said, a similar exercise will need to be undertaken for societal challenges and forms of evidence that are quite different from those described here. For example, the Intergovernmental Panel on Climate Change (IPCC) has helped a great deal with global coordination in their area of focus, and with spurring new approaches to modeling over long time horizons. However, the IPCC may also benefit from complementing these approaches with post-hoc evaluations of climate-change response options.

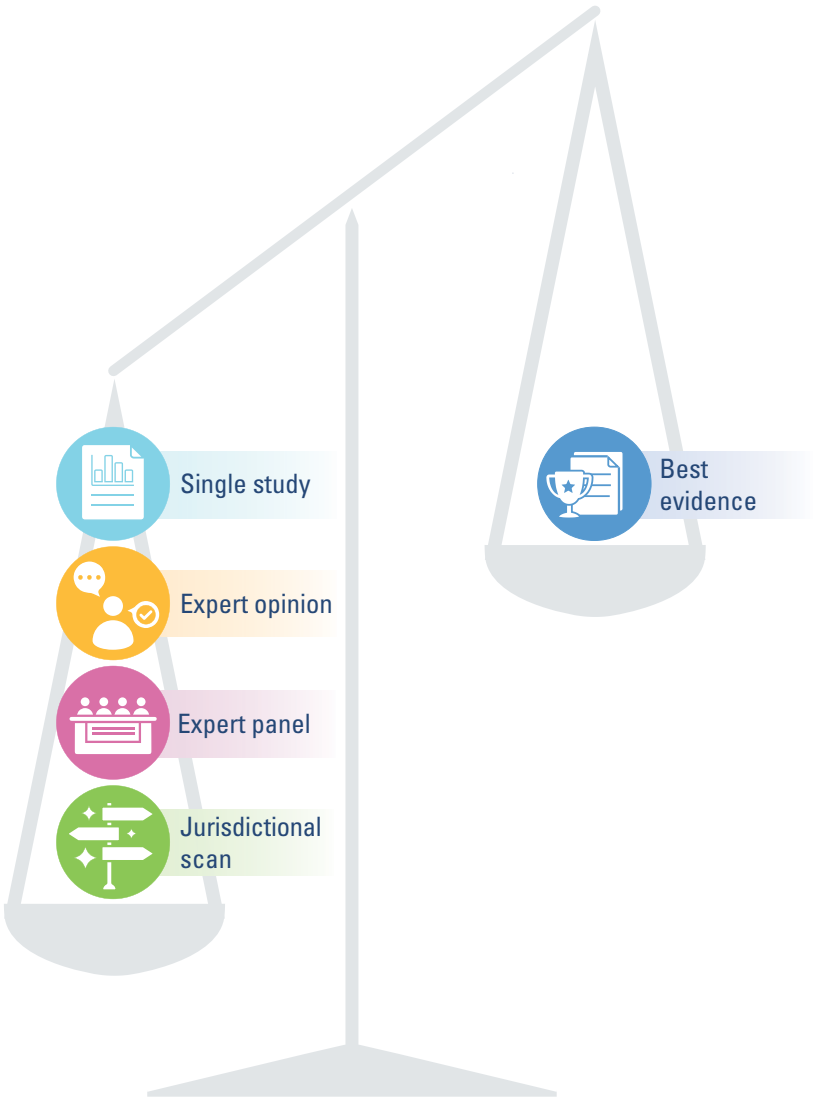
Pre-COVID weaknesses in the health-research system	Examples of weaknesses that became more apparent through the COVID-19 evidence response	Examples of efforts to address weaknesses through the COVID-19 evidence response
Lack of **global coordination** of evidence communities, with each ideally addressing a globally prioritized challenge using systematic and transparent methods and a full array of data sources *(e.g., study registries, regulatory agencies, and administrative databases)*	• Many topics prioritized by COVID-END's global horizon-scanning panel were never addressed by one or more 'best' evidence syntheses • Low signal-to-noise ratio: nearly 11,000 evidence syntheses about COVID-19 were reduceable to roughly 600 'best' evidence syntheses in the COVID-END inventory (as of 7 November 2021) based on four criteria: addressing a unique decision-relevant question, recency of the search for evidence, quality of the synthesis, and availability of a GRADE evidence profile	• COVID-END engaged 55 leading evidence-synthesis, guideline-development and technology-assessment groups, as well as citizen partners and evidence intermediaries, in efforts to reduce duplication and enhance coordination • PROSPERO encouraged those registering a protocol for a COVID-19 evidence synthesis to search for already registered protocols and to pick a new topic if duplication was likely (although 138 teams proceeded with a topic already registered by one of 57 other teams, including 14 addressing hydroxychloroquine and seven addressing tocilizumab) • GloPID-R (Global Research Collaboration for Infectious Disease Preparedness) engaged leading research-funding organizations in coordinating their rapid funding of primary research about COVID-19
Lack of focus of evidence communities on maintaining **living evidence syntheses** that examine all interventions addressing a prioritized challenge *(e.g., a network meta-analysis rather than pairwise comparisons only)*	• Only 13% of COVID-19 evidence syntheses self-identified as a living evidence synthesis (versus 52% in the COVID-END inventory where 'living' status was a criterion used to identify 'best' evidence syntheses) and more than two thirds addressed clinical management (rather than public-health measures, health-system arrangements, and economic and social responses) • Only 21% of living COVID-19 evidence syntheses had one update (after the first publication), 8% had two, and 13% had two or more, while the mean and median time between searches for syntheses with updates was 49 and 31 days, respectively • Many COVID-19 evidence syntheses addressed single drug treatments, so the COVID-END inventory transitioned to relying primarily on COVID-NMA and others looking across all drug treatments (and to including only syntheses of prognostic studies that include all available prognostic factors)	• Four evidence communities maintained high-quality living meta-analyses of all drug treatments, with one (COVID-NMA) supporting weekly updates of risk-of-bias assessments and GRADE certainty assessments

Lack of focus of evidence communities on identifying **harms** arising from interventions as well as benefits *(and more generally including a broader array of study designs and types of data)*	• Then-existing studies and syntheses made it difficult to understand what to make of reports about blood clots being experienced by select vaccine recipients	• A COVID-END team conducted a systematic review to complete a causality assessment of thrombotic thrombocytopenia that is temporally related to vaccine administration
Lack of sharing of **individual participant data** and its use to examine how findings vary by type of participant, setting or other factors, and hence how interventions can be better personalized or contextualized	• Many reports documented the lack of sharing of individual participant data (e.g., one review of 140 studies early in the pandemic found that data were shared from only one study – see bit.ly/31WQUxM)	• The COVID-19 Knowledge Accelerator advanced the methods needed to share computable expressions of evidence and guidance across platforms, and Vivli extended its platform to enable the sharing of COVID-19 trials data
Lack of **inclusion** in evidence communities of representatives from all relevant evidence groups *(e.g., researchers conducting primary studies like trials, evidence synthesizers and guideline developers)*, all relevant types of decision-makers, and all relevant types of evidence intermediaries	• Many reports described how citizens were less involved in COVID-19 research than they had been in other types of research before the pandemic, as well as about plain-language summaries of evidence syntheses not being available early in the pandemic (e.g., bit.ly/3kwCHhr)	• The National COVID-19 Clinical Evidence Task Force involved many health professionals (and their associations) and patients in their living guidelines, and they worked in partnership with evidence communities maintaining living network meta-analyses • Many groups engaged in modeling to help choose among available options (e.g., lockdowns) based on available evidence and expert opinion, and in some cases the context provided by decision-makers • Many groups prepared contextualized rapid syntheses at the request of decision-makers (with citizen partners in the case of many COVID-END rapid syntheses)
Lack of use by evidence communities of a range of **new approaches** to become more efficient and timely in their work *(e.g., machine learning and crowd-sourcing contributions to their work)*	• More than 18,000 studies had been uploaded to just one preprint server (medRxiv) by July 2021, dramatically shortening the time to publication (while having uncertain harms due to the lack of peer review) • Many use cases for machine-learning approaches in COVID-19 responses were identified in a medium-quality scoping review of 183 reports (bit.ly/3D7bTeV), but were not widely used early in the pandemic	• L*VE (Living Overview of Evidence) used machine learning to maintain a repository of primary studies and evidence syntheses, and the EPPI-Centre used machine learning to maintain a living evidence map
Lack of **reporting** about the gaps in and quality and transparency of primary studies (including conflicts of interest) as part of a feedback loop meant to support learning and improvement – for more details, see box 1 in this paper: (17)	• The results of many primary studies have been made available through media releases instead of through full research reports that can be critically appraised • Many reports noted that primary studies were found to have an intermediate to high risk of bias (e.g., 81% of the 713 articles including original patient data from a pool of 10,516 COVID-19 articles – see bit.ly/3HiI90X) and to have been retracted because of scientific misconduct • COVID-END prepared reports about evidence syntheses' lack of currency (91% and 61% in the full database and inventory of 'best' evidence syntheses, respectively, were based on searches completed more than 180 days earlier), medium or low quality (75% and 55%, respectively), and lack of an evidence profile (81% and 42%, respectively), as well as how rapid syntheses were more likely to be low quality than full syntheses (43% compared to 13%)	• RECOVERY (recoverytrial.net) and WHO COVID Solidarity Therapeutics Trial provided platforms for ultra-rapid, high-quality, multi-country trials of COVID-19 drug treatments • COVID-19 Evidence Alerts profiled quality-rated primary studies

4.13 Weaknesses in many COVID-19 evidence-support systems

The COVID-19 pandemic has been a global crisis marked by the need for rapid-fire decision-making by high-level government authorities over several 'waves', and by both significant uncertainty and a quickly evolving (and often indirect) evidence base. In many jurisdictions, evidence appeared to play a more visible role in government policymaking during the COVID-19 pandemic than it has in many decades. That said, misinformation flourished, and citizens and other stakeholders struggled to understand why the evidence changed over time. 'Other things' than best evidence often had greater visibility than best evidence, and some forms of evidence often had greater visibility than others. We addressed misinformation in **section 4.11** and we provided additional context for the terms used here in **sections 4.8** ('other things' than best evidence), **4.2** (forms of evidence) and **4.5** (distinguishing high- from low-quality evidence).

'Other things' than best evidence that were more typically encountered by COVID-19 decision-makers



Single study *(often as a preprint)*
Risk of 'hubcap chasing'* unless each study was quality assessed and then either considered as local (national or sub-national) evidence or put in the context of a living (global) evidence synthesis

Expert opinion
Risk of 'squeaky wheel getting the grease' unless the expert was asked to share the quality-assessed evidence syntheses on which their opinion was based, or to focus on what specific evidence syntheses mean for a given jurisdiction

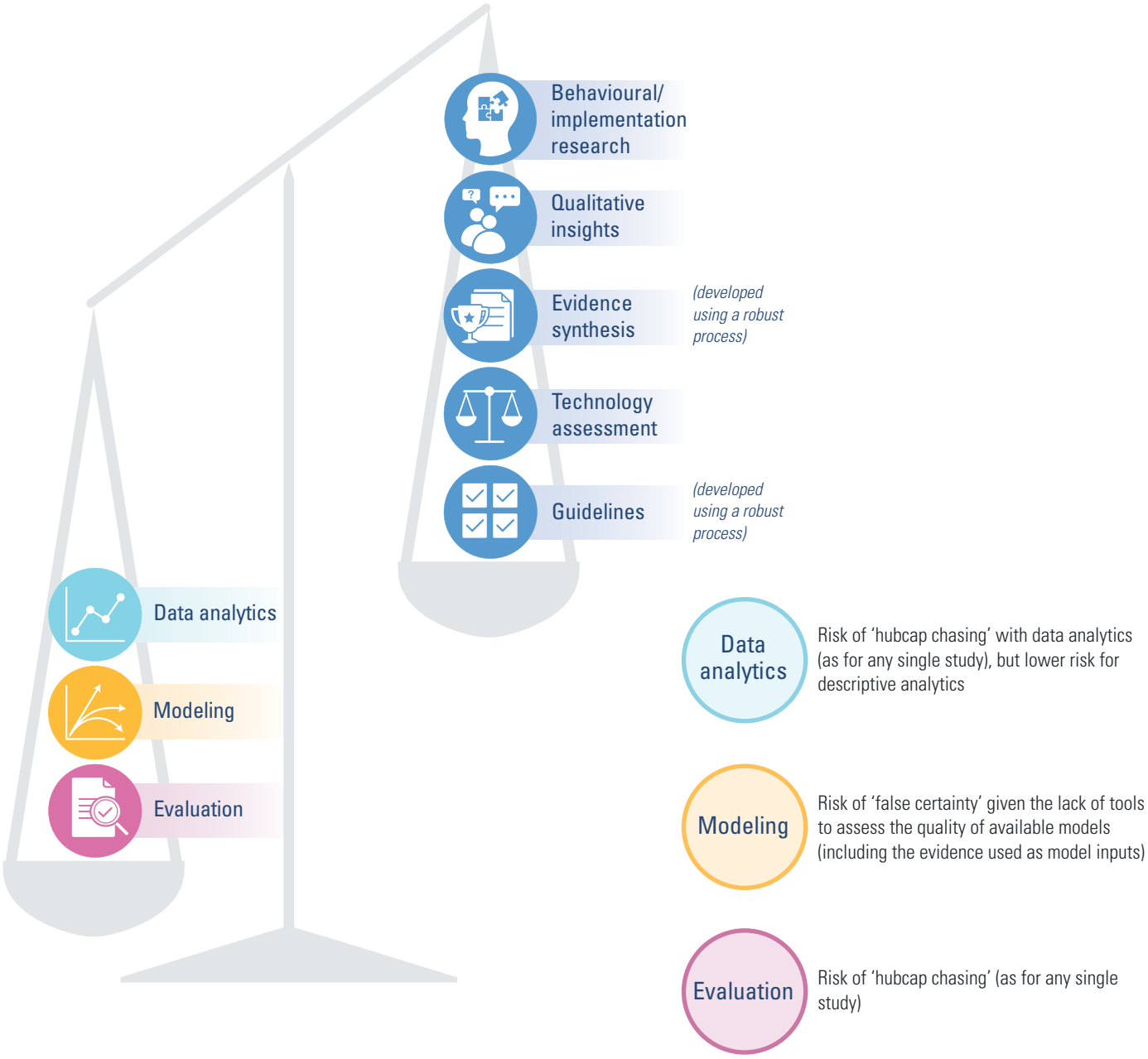
Expert panel
Risk of GOBSATT (or 'good old boys sitting around the table') unless the panel members were asked to share their evidence (as above) or were supported by a robust guideline-development process

Jurisdictional scan
Risk of 'groupthink' unless jurisdictions shared their supporting evidence or plans for generating it

Best evidence
In a given national (or sub-national) context: national (or sub-national) evidence drawn from the best available studies (i.e., what has been learned in that context) and global evidence drawn from the best available evidence syntheses (i.e., what has been learned from around the world, including how it varies by groups and contexts)

** As noted in **section 4.8**, we use the term 'hubcap chasing' (i.e., dogs repeatedly barking at and chasing cars) as a metaphor for sharing and commenting upon each new study that captures one's attention.*

Forms of evidence that were more typically encountered by COVID-19 decision-makers



Leaders in any jurisdiction can use the Evidence Commission report to systematize and broaden beyond health the aspects of the evidence response to COVID-19 that went well and to address the many aspects that did not go well. As part of systematizing what went well, these leaders will need to transition from the COVID-19-era focus on speed and as much quality as possible ('quick and clean enough') to a balance among speed, quality (e.g., waiting for evidence that is just around the corner), and sustainability (e.g., normal working hours and other work not put on hold).

4.14 Features of an ideal national evidence infrastructure

Every country has a national evidence infrastructure that includes many evidence-related structures and processes. Within this national evidence infrastructure, we distinguish the evidence-support system, the evidence-implementation system, and the research system. Giving much greater attention to the evidence-support system, and ongoing attention to the evidence-implementation system, will be key to future efforts to use evidence in addressing societal challenges.

Evidence is something that decision-makers can use, while research is something that researchers do. When decision-makers ask a question, particularly government policymakers and organizational leaders, they need to be supported in a timely way in using the evidence that already exists. Decision-makers, particularly professionals and citizens, need to be supported to implement the changes that robust evidence demonstrates are needed. Meanwhile, researchers need to be enabled to invent new products and services, to develop new ways of thinking, and to critique the status quo. They also need to be encouraged to engage more actively with decision-makers to ensure relevance and applicability, to use technology more effectively to make research processes more efficient, to report their findings more transparently and without 'spin,' and to create versions of the evidence they produce that can be accessed, understood and made actionable by decision-makers. The evidence emerging from their research that is 'ready for prime time' can then be drawn into the evidence-support and evidence-implementation systems.

		Enabler	Complement
Evidence-support system	Grounded in an understanding of a national (or sub-national) context (including time constraints), demand-driven, and focused on contextualizing the evidence for a given decision in an equity-sensitive way Examples of infrastructure: • evidence-support coordination office (for all of government, with or without additional offices in key departments or ministries) • evidence units with expertise in each of eight forms of evidence (e.g., behavioural-insights unit) • processes to elicit and prioritize evidence needs, find and package evidence that meets these needs within set time constraints (and build additional evidence as part of ongoing evaluations), build capacity for evidence use (e.g., evidence-use workshops and handbook), prompt evidence use (e.g., cabinet submission checklist), and document evidence use (e.g., evidence-use metrics) *While such infrastructure is most relevant to government policymakers and the leaders of very large organizations, similar types of infrastructure can be tailored to the leaders of smaller organizations as well as professionals and citizens*	Enabled by: • domestic evidence intermediaries • evidence-related global public goods (e.g., global standards and open-access publications of evidence syntheses) from Cochrane, Campbell and others • technical assistance from the UN and other multilateral organizations, including their country, regional and global offices	Complemented by: • foresight initiatives to anticipate future evidence needs • innovation hubs to invent new products and services, evaluate them, and scale those that can add value through markets or public procurement

		Enabler	Complement
Evidence-implementation* system	Grounded in an understanding of evidence-related processes, driven by a mix of demand and supply considerations, and focused on cycles of synthesizing evidence, developing recommendations, disseminating them to decision-makers, actively supporting their implementation, evaluating their impacts, and incorporating lessons learned in the next cycle (18) Examples of infrastructure: • evidence-synthesis and guideline units • evidence-implementation units to prioritize what to implement, identify barriers and facilitators to implementation, and design strategies that address barriers and leverage facilitators • processes to build evidence into existing workflows (e.g., electronic client records, digital decision-support systems, web portals, and quality-improvement initiatives) and share it across them *While such infrastructure is most relevant to professionals and citizens, similar types of infrastructure can be tailored to government policymakers and organizational leaders*	Enabled by similar things as above	Complemented by government policymakers and organizational leaders using available levers to support implementation (e.g., adding recommended products and services to a benefits package, and mandating public reporting of an indicator capturing adherence to a recommended action)
Research system	Grounded in an understanding of disciplinary perspectives and research methods, driven by supply-side considerations like curiosity, and focused on conducting research that may or may not aim to contribute to the evidence taken up in the evidence-support and evidence-implementation systems (19) Examples of infrastructure: • university departments and units • processes to reward activities (e.g., peer-reviewed grants and publications), which could be expanded to activities with a greater likelihood of achieving impacts (e.g., engagement with and responsiveness to decision-makers) *Such infrastructure is most relevant to researchers*	Enabled by research-related global public goods (e.g., open-science initiatives)	Complemented by government policymakers and organizational leaders using available levers to reward certain activities (e.g., institution-assessment exercises like the UK's Research Excellence Framework)

**We use the term evidence-implementation system to distinguish it from the evidence-support system. Some recent descriptions of what we mean by an evidence-implementation system have called this an evidence ecosystem.(18) We have avoided this term both because it confuses those who are used to the literal meaning of an ecosystem and because it does not capture this system's focus on implementation. If we were to use the term evidence ecosystem, we would likely apply it to a combination of the evidence-support system and the evidence-implementation system.*

Building on the first row above, an evidence-support system would ideally have the following features:

- supports decision-making by government policymakers, as well as by organizational leaders, professionals and citizens, with the best evidence and in ways that are:
  - informed by a good understanding of their context – including where and how decisions are made, the time constraints under which decisions are made, and the existing system arrangements that determine whether the right products and services get to those who need them – and of their capacities, opportunities and motivation to use evidence in decision-making
  - responsive to their decision-related needs, time constraints, and preferences for product and process formats
  - reflective of a commitment to matching the best evidence to the question asked and to working through what the evidence means for a given decision (i.e., to contextualizing the evidence), including how this may vary by groups and contexts (i.e., to bringing an equity lens to the evidence and to how it is viewed)
  - delivered with judgement, humility and empathy and with appropriate attention to identifying and managing conflicts of interest
- enabled in systematic and transparent ways both by those within government and through strategic partnerships with evidence intermediaries and producers outside government, such as domestic evidence intermediaries and purveyors of global public goods and technical assistance
- complemented by those operating in two parts of what the UN calls its 'quintet of change,' namely strategic foresight and innovations.(20)

The three other parts of the quintet of change – data analytics, behavioural/implementation research, and evaluation ('performance and results orientation') – are captured in our eight forms of evidence.

Some governments have chosen to pass legislation that formalizes aspects of the evidence-support system. In the US, the bipartisan Commission on Evidence-based Policymaking (21) developed recommendations that informed the Evidence Act. Follow-up memos from the president and the Congressional Budget Office helped to support the implementation of the act. These efforts share with the Evidence Commission a focus on all types of societal challenges, but diverge in their focus on just one type of decision-maker (government policymakers, in this case in the US federal government), on just two forms of evidence (data analytics and evaluation), and on building new evidence and not also on making better use of the stock of existing evidence (such as through evidence syntheses). Some parts of the UN system have chosen to pass resolutions about strengthening evidence-support systems. In the Eastern Mediterranean region, WHO's regional committee passed such a resolution for the health sector.(22)

4.15 Global-commission reports by form of evidence

Only one of 70 global commission reports published since January 2016, in describing their commissioners singled out expertise in any of the eight forms of evidence that decision-makers typically encounter.

When commission reports explicitly reported in their methods section that they drew on any of these forms of evidence in their own work, modeling was the most frequent form (13 reports) and evidence synthesis (6) and technology assessment / cost-effectiveness analysis (5) were the next most frequent. Complementing this analysis of methods sections, an analysis of reference lists found:

- 64 of 70 reports had a reference list
- only 32 of these 64 reports had at least one citation of an evidence synthesis
- only 3% of citations (526 of 17,605) appeared to be evidence syntheses based on their titles
- the mean and median number of citations of evidence syntheses were 8.2 and one per report, respectively.

We also analyzed the citation list for the Global Sustainable Development Report 2019, which was prepared by an independent group of scientists appointed by the UN Secretary-General and which, accordingly, one might expect to be a positive outlier.(20) However, in this report only 1.8% of citations (17 of 941) appeared to be evidence syntheses based on their titles. When evidence syntheses were cited, it wasn't clear whether quality and recency of search played a role in selecting them. For example, three of the cited evidence syntheses addressed the specific topic of health-worker recruitment and retention, yet there are hundreds of syntheses available on this topic through Health Systems Evidence. We focus on evidence syntheses because – as we note in **section 4.2** – they use a systematic process of identifying, selecting, appraising and synthesizing the findings from all studies that have addressed the same question to arrive at an overall understanding of what is known, including how this may vary by groups and contexts.

At most one of the reports made any one of these forms of evidence the explicit focus of their recommendations. As we return to in **section 7.1**, many reports made general recommendations about data collection and sharing, but they did not make specific recommendations about harnessing data analytics to support decision-making.

Forms of evidence		Number of commission reports
Basis for describing the expertise of members of the commission (not including their individual bios)	Technology assessment / cost-effectiveness analysis	1
	All other forms of evidence	0
	Not explicitly reported	69
Source of evidence drawn upon	Modeling	13
	Evidence synthesis	6
	Technology assessment / cost-effectiveness analysis	5
	Data analytics	3
	Evaluation	2
	Guidelines	2
	Behavioural/implementation research	1
	Qualitative insights	1
	Not explicitly reported	49
Focus of recommendations	Modeling	1
	Evaluation	1
	Qualitative insights	1
	Technology assessment / cost-effectiveness analysis	1
	Guidelines	1
	All other forms of evidence	0
	Not explicitly reported	66

4.16 Annex to section 4.5 – Examples of quality-assessment tools

Form of evidence	Examples of quality-assessment tools
Types of evidence for which quality-assessment tools exist	
Data analytics	ROBINS-I (riskofbias.info) for observational studies, such as those that examine associations between select factors (including interventions) and select outcomes, where there is a risk of bias from: • confounding (where the observed relationship between a factor and an outcome, differs from the true relationship because of one or more additional factors that are not accounted for) • selection of participants into the study • classification of intervention(s) • deviations from intended intervention • missing data • measurement of outcomes • selection of the reported result
Evaluation	Risk of Bias (RoB) 2 (riskofbias.info) for randomized-controlled trials, where the risk of confounding is less, but where there is a risk of bias from some (albeit fewer) of the same sources as above: • randomization process • deviations from the intended interventions • missing (outcome) data • measurement of outcomes • selection of the reported result
Behavioural/ implementation research	See other rows for the relevant types of studies or syntheses
Qualitative research	JBI critical appraisal checklist for qualitative research (bit.ly/31Lsib1), where very different considerations come into play, such as: • congruity between the research methodology and the research question, data-collection methods, data representation and analysis, and results interpretation, as well as between the stated philosophical perspective and the methodology • reflexiveness on the part of the researcher, such as statements locating the researcher culturally and theoretically, and addressing the researcher's influence on the research and vice versa • representation of study participants and their voices • flow of conclusions from the analysis and interpretation of the data
Evidence synthesis	See above for the relevant types of studies considered in the evidence synthesis
	A MeaSurement Tool to Assess systematic Reviews (AMSTAR; amstar.ca) for the quality of the evidence synthesis, where the risk of bias can arise from: • identification of all potentially relevant studies through a comprehensive search of both published and grey literature and without language restrictions • selection of all studies addressing the research question using explicit criteria about study designs and about participants, interventions/factors, comparisons and outcomes, and with at least two reviewers applying the criteria • quality appraisal of and data extraction from all included studies • synthesis of findings from all included studies Note that there are two versions of AMSTAR: 1) the original version that can be applied across all types of syntheses, albeit with some criteria removed from both the numerator and denominator; 2) a second version of AMSTAR that is more specifically relevant to syntheses of randomized-controlled trials
	Grading of Recommendations, Assessment, Development and Evaluations (GRADE; bit.ly/3C9pMrx) for the certainty of evidence for the outcomes of an intervention, with: • certainty rated down because of risk of bias (with evidence from randomized-controlled trials starting at high certainty and evidence from observational studies starting at low quality and then being adjusted based on RoB2 or ROBINS-I), imprecision (e.g., one or two small studies), inconsistency (e.g., two studies showing very different findings), indirectness (e.g., surrogate measures used or atypical settings studied), and publication bias (e.g., more common with observational studies because of the lack of study registries or with industry-funded studies because of the commercial incentive to publish positive studies) • certainty rated up for large magnitude of effect, dose-response gradient, and when all residual confounding would decrease the magnitude of effect
	GRADE CERQual (cerqual.org) for the certainty of evidence for the qualitative representation of a phenomenon of interest, with: • certainty rated down because of concerns about methodological limitations (because problems in the way studies were designed or reported were identified using a critical-appraisal tool like the JBI one above), relevance (because the context in which the primary studies were conducted are substantively different from the context of the synthesis question), coherence (because some of the data contradict the findings or are ambiguous), and adequacy (because the data are not sufficiently rich or only come from a small number of studies or participants)

Technology assessment / cost-effectiveness analysis	International Network of Agencies for Health Technology Assessment (INAHTA) checklist (bit.ly/2YJVMVK) for the quality of technology assessments, with two of the 14 questions addressing the approach to synthesizing the evidence (with prompts similar to AMSTAR) and another question addressing whether the assessment was contextualized through an accompanying cost-effectiveness analysis (with local – meaning national or sub-national – costing data), and consideration of local legal, ethical and social implications
	Drummond checklist of cost-effectiveness analyses (bit.ly/3FbnB8R), and for economic evaluations more generally, with questions about study design, data collection, and the analysis and interpretation of results
	Philips checklist for cost-effectiveness analyses that include a decision-analytic modeling component (bit.ly/3FcWBGc) with questions about the structure of the model (e.g., explicit rationale, justified assumptions and appropriate time horizon), the data used (e.g., baseline probabilities from observational studies, treatment effects from randomized-controlled trials, and assessments of four types of uncertainty, namely the structure of the model, the methodological steps followed, the heterogeneity in the population studied, and the parameters used), and the consistency (internal and external) – there is also the complementary TRUST tool to assess uncertainties in decision-analytic models (bit.ly/3quFSKp)
Guidelines	AGREE II tool (bit.ly/30qyFAb) for assessing the development, reporting and evaluation (or quality appraisal) of guidelines, which uses 23 items grouped into six domains, each of which is scored independently: • scope and purpose described • stakeholder (citizen/patient and professional) involvement • rigour of development (with evidence syntheses used as an input, a robust recommendations-development process, and recommendations linked to the supporting evidence) • clarity of presentation • applicability • editorial independence (in relation to funder and panel members' conflicts of interest)
	Grading of Recommendations, Assessment, Development and Evaluations (GRADE; bit.ly/3C9pMrx) for assessing the strength of recommendations, which uses four key considerations: • balance between desirable and undesirable outcomes (trade-offs), taking into account best estimates of the magnitude of effects on desirable and undesirable outcomes, and the importance of those outcomes (estimated typical values and preferences) • confidence in the magnitude of estimates of effects of the interventions on important outcomes (see GRADE in a previous row) • confidence in values and preferences and their variability resource use
Types of evidence for which quality-assessment tools don't yet exist	
Modeling	No widely accepted tool yet exists for most types of models, however, there are some general questions that can be asked about models (much like those listed as part of the Philips checklist above), such as: • structure of the model (e.g., explicit rationale, justified assumptions, and appropriate time horizon) • data used (e.g., baseline probabilities from observational studies, intervention effects from a range of sources*, and assessments of four types of uncertainty, namely the structure of the model, the methodological steps followed, the heterogeneity in the population studied, and the parameters used) • consistency (internal and external) • availability of the software or tool so that it can be assessed by others *One of the challenges with COVID-19 was that study designs typically used to capture intervention effects, such as randomized-controlled trials, were ethically or logistically difficult and/or took time to complete, so other study designs needed to be used and expert opinion needed to be sought (and there are approaches that enable this to be done in a way that is systematic and transparent, such as SHELF – see bit.ly/30nteC4)
Approaches used with certain types of evidence for which quality-assessment tools don't yet exist	
Artificial intelligence	No widely accepted tool yet exists

4.17 References

1. Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. *Science Advances* 2021; 7(21).
2. Minx JC, Haddaway NR, Ebi KL. Planetary health as a laboratory for enhanced evidence synthesis. *Lancet Planet Health* 2019; 3(11): e443-445.
3. Yusuf M, Atal I, Li J, et al. Reporting quality of studies using machine learning models for medical diagnosis: A systematic review. *BMJ Open* 2020; 10(e034568).
4. Lavis J, Oxman A, Souza N, Lewin S, Gruen R, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). 9. Assessing the applicability of the findings of a systematic review. *Health Research Policy and Systems* 2009; 7(S9).
5. Connolly P, Keenan C, Urbanska K. The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980-2016. *Educational Research* 2018; 60(3): 276-291.
6. Page M, Shamseer L, Altman D, Tetzlaff J, Sampson M, Tricco AC. Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLoS Medicine* 2016; 13(5): e1002028.
7. Phillips D, Coffey C, Tsoli S, Stevenson J, Waddington H, Eyers J. A map of evidence maps relating to sustainable development in low- and middle-income countries: Evidence gap map report. London: CEDIL Pre-Inception Paper; 2017.
8. Thyer B. A bibliography of randomized controlled experiments in social work (1949-2013: Solvitur Ambulando. *Research on Social Work Practice* 2015; 25(7): 753-793.
9. Mavergames C, Elliott J. Living systematic reviews: Towards real-time evidence for health-care decision-making. London: BMJ Best Practice. https://bestpractice.bmj.com/info/toolkit/discuss-ebm/living-systematic-reviews-towards-real-time-evidence-for-health-care-decision-making/ (accessed 7 November 2021).
10. Kirkland D. No small matters: Reimagining the use of research evidence from a racial justice perspective. New York: William T. Grant Foundation; 2019.
11. Vadehra E. We need to reimagine the modern think tank. 2021. https://ssir.org/articles/entry/we_need_to_reimagine_the_modern_think_tank# (accessed 12 November 2021).
12. Broadband Commission for Sustainable Development. Balancing act: Countering digital disinformation while respecting freedom of expression. Geneva and Paris: International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization; 2020.
13. Walter N, Murphy S. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs* 2018; 85(3): 423-441.
14. Douthat R. How I became extremely open-minded. *New York Times* 2021; 6 November 2021.
15. Boutron I, Créquit P, Williams H, Meerpohl J, Craig JC, Ravaud P. Future of evidence ecosystem series: 1. Introduction: Evidence synthesis ecosystem needs dramatic change. *Journal of Clinical Epidemiology* 2020; 123: 135-142.
16. Créquit P, Boutron I, Meerpohl J, Williams HC, Craig J, Ravaud P. Future of evidence ecosystem series: 2. Current opportunities and need for better tools and methods. *Journal of Clinical Epidemiology* 2020; 123: 143-152.
17. Ravaud P, Créquit P, Williams HC, Meerpohl J, Craig JC, Boutron I. Future of evidence ecosystem series: 3. From an evidence synthesis ecosystem to an evidence ecosystem. *Journal of Clinical Epidemiology* 2020; 123: 153-161.
18. Vandvik P, Brandt L. Future of evidence ecosystem series: Evidence ecosystems and learning health systems - Why bother? *Journal of Clinical Epidemiology* 2020; 123: 166-170.
19. Pang T, Sadana R, Hanney S, Bhutta ZA, Hyder AA, Simon J. Knowledge for better health: A conceptual framework and foundation for health research systems. *Bulletin of the World Health Organization* 2003; 81(11): 815-20.
20. United Nations. UN 2.0: Quintet of change. New York: United Nations; 2021.
21. Commission on Evidence-Based Policymaking. The promise of evidence-based policymaking. Washington: United States Government Printing Office; 2017.
22. World Health Organization Regional Office for the Eastern Mediterranean. Resolution on developing national institutional capacity for evidence-informed policy-making for health. Cairo: World Health Organization; 2019.
23. Independent Group of Scientists appointed by the Secretary-General. Global sustainable development report 2019: The future is now – Science for achieving sustainable development. New York: United Nations; 2019.