



第四章 研究、综合与指南——证据供给

4.1 决策中常见的证据形式	44	4.10 原住民权益及其认知方式	60
4.2 常见证据形式的定义	45	4.11 错误信息和信息疫情	62
4.3 匹配决策相关问题与证据形式	47	4.12 卫生研究系统的局限性	65
4.4 本土证据和全球证据的相互作用	48	4.13 众多COVID-19证据支持系统的不足	67
4.5 高质量证据和低质量证据的区别	50	4.14 国家证据基础设施的理想特征	69
4.6 证据综合的覆盖面、质量和新近度	52	4.15 以证据形式区分的全球委员会报告	72
4.7 动态证据产品	55	4.16 质量评价工具的示例（第4.5节的附录）	73
4.8 最佳证据与其他证据的对比 (以及如何从其他证据中获得最大收益)	57	4.17 参考文献	75
4.9 形成如何看待证据的环境	59		

本章是本报告核心问题三个章节中的第三章，即在应对社会挑战时，所有的决策者如何系统化地使用证据。在此，我们重点探讨研究、综合与指南，即证据供给。第二章关注社会挑战的本质；第三章关注决策和决策者，或对证据的需求。

版权所有©2022麦克马斯特大学。保留所有版权。本报告采用创作共享署名4.0国际许可证授权。未经出版商事先书面许可，不得以任何形式改编本报告的任何部分。

本报告和其中包含的信息仅供参考和公共利益所用。虽然秘书处和委员们已努力确保撰写本报告时信息是最新和准确的，且按原样发布，无任何形式的明示或暗示担保。本报告中包含的信息并非用于替代财务、法律或医疗建议。

麦克马斯特大学、证据委员会秘书处、委员们和出版商对因使用本报告中包含的信息而直接或间接造成或据称造成的损失或损害不承担责任或义务。麦克马斯特大学、秘书处、委员们和出版商明确否认因使用或应用本报告所含信息而产生的任何责任。

本报告的出版商是麦克马斯特卫生论坛（McMaster Health Forum）（地址：加拿大安大略省汉密尔顿市主路西1280号米尔斯纪念图书馆417号，1280 Main St. West, MML-417, Hamilton, ON, Canada L8S 4L6）。麦克马斯特卫生论坛代表证据委员会，欢迎读者对本报告提出任何反馈和影响报告建议途径的意见。请将您的意见或建议发送至邮箱 evidencecommission@mcmaster.ca。

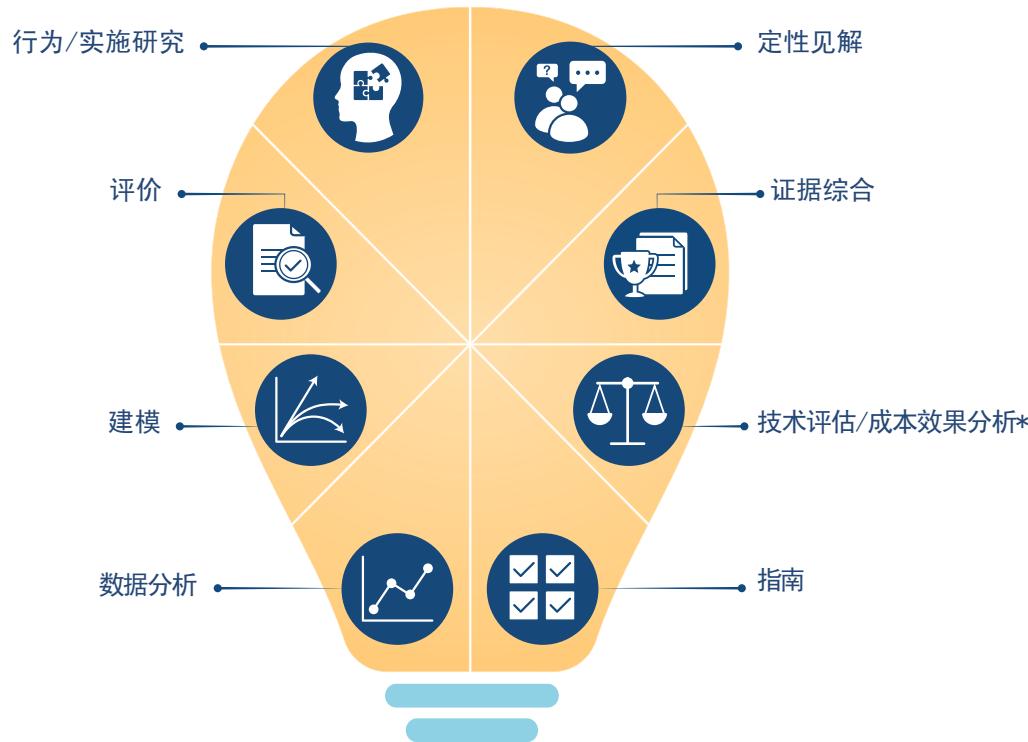
本报告的引用：

应对社会挑战的全球证据委员会. 第四章：研究、综合与指南——证据供给. 证据委员会报告：为决策者、证据中介和以影响力为导向的证据生产者敲响警钟并指明前进道路 [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]. 汉密尔顿：麦克马斯特卫生论坛，2022年；p. 43–75.

ISBN 978-1-927565-39-1 (Online)
ISBN 978-1-927565-33-9 (Print)

4.1 决策中常见的证据形式

决策过程中通常会遇到八种不同形式的证据，这些形式相互关联。比如，随机对照试验的评价也有可能包含基于数据分析、定性见解和成本效果分析的证据。同样，案例研究可以采用关于经验和偏好的定性见解，也可以使用来自数据分析、建模和评价的定量证据。



*我们将技术评估和成本效果分析归为一类，因为它们通常来源于相同的证据组，应用于相同类型的产品或服务，而且成本效果分析几乎总是技术评估的关键因素。我们认识到，某些形式的证据生产者相比于生产证据产品更强调其过程，但是许多没有参与过任何相关过程的决策者仍会面对这些形式的证据。

本章节题目中提到的“研究”可以产生许多形式的证据（例如，评价，行为研究，定性研究和其他形式的“原始”研究）。本章节题目中提到的“综合”本身就是一种证据形式，有时也被称为“二次”研究。本章节题目中提到的指南同样也是一种证据形式，正如我们在第4.4节所讨论的一样，技术评估也可以包括推荐意见。

我们使用“证据”一词作为“研究证据”的简称，因为我们意识到还有许多其他形式的证据（例如，来源于个人生活经验的证据和法庭上考虑的证据），并且证据是影响决策的众多因素之一。我们在第4.2节对每一条术语进行了定义，并展示了每种形式的证据如何影响决策过程。我们在第4.3节中描述了相反的情况，即决策过程中的各步骤如何与证据形式相联系。

4.2 常见证据形式的定义

我们在下面提供了每种证据形式的简单定义。我们对其他研究者的定义做了许多改编，目的是为了更清楚地区分这八种证据形式，同时也展现它们之间的相关性。同时，我们说明了每种形式的证据是如何与决策过程中的四个步骤相互联系的。



本节与第4.3节（介绍决策过程各步骤与证据形式的关系），均以在第3.1节中首次介绍的决策问题清单为基础。

证据形式	定义	可发挥最大价值的步骤			
数据分析 	对原始数据进行系统分析，以得出关于该信息的结论	1			4
建模 	在虚拟环境中使用数学方程模拟现实世界情景（即，如果我们不干预可能会发生什么）和选择（即，如果我们干预会发生什么）	1	2		
评价 	以学习或决策为目的，对某一方案的实施（监测）和效果（评估）进行的系统评价				4
行为/实施研究 	研究如何在个人、专家人员、组织及政府层面将有效方法系统地应用于日常工作中（实施研究） 系统地考核人们（公民和专业人士）在做什么，是什么促使他们这样做，以及什么可以维持或改变他们的行为（行为研究）			3	

定性见解 	数据研究（通常是非数值的），即通过访谈、焦点小组、开放式问卷、一手观察、参与观察、自然环境中的记录、文件和人工制品中获取，以了解个人和群体对于问题、方案、实施注意事项（阻碍因素、促进因素和策略）以及衡量标准的看法和经验				
证据综合 	识别、筛选、评价和综合所有针对同一问题研究结果的系统过程，以全面了解现有知识，包括结果可能如何因人群（如种族化社区）和环境（如社会经济地位低下的社区）而有所不同				
技术评估/成本效果分析 	<p>对于某种“技术”（如产品或服务）的全面评估，包括安全性、有效性以及经济、社会和伦理影响（技术评估），证据综合通常有助于评估有效性</p> <p>比较两种或多种选择的相对结果（有效性）和成本，同样，证据综合通常有助于评估有效性</p>				
指南 	系统制订的声明用以推荐特定的行为准则。该声明通常针对公民和专业人士，有时针对组织和政府，并且通过一部或多部证据综合评估有效性、价值观和偏好以及其他因素				

*在此步骤中发挥最大价值，但在其他步骤中也可以发挥作用

请注意，从任何形式的证据或任何形式的证据组合中衍生出来的简报、信息图表、简明摘要和其他文件，都可以用来为不同类型的决策者打包传递关键信息。这种“衍生证据产品”可作为传播和实施方案应用于此类决策者，并在所有步骤中增加其价值。

4.3 匹配决策相关问题与证据形式

在第4.2节中，我们描述了证据形式与决策过程中各步骤间的相关性，在此节，我们将决策过程中的每个步骤与证据形式联系起来，并举例说明。

证据综合通过对解决一个相似问题的所有研究中已知和未知信息进行总结，可以帮助回答几乎所有相关问题。对于利弊相关问题，证据综合无论是在方案选择还是在实施策略中都至关重要。我们在第4.4节中详述了为什么证据综合是回答多种类型问题的最佳起点。



步骤	相关问题	有帮助的证据形式示例
1	指标 —— 问题有多大? 比较 —— 问题是否越来越严重或者比他处更严重? 框架 —— 不同的人如何描述或体会这个问题及其原因?	数据分析 数据分析（例如，使用行政数据库或社区调查） 定性研究（例如，使用访谈或焦点小组）
2	利 —— 它可能带来什么好处? 弊 —— 可能出现什么问题? 成本效果 —— 在相同的投入下，某个方案能否获得更佳效果? 适应性 —— 我们能否对在其他地方行之有效的内容进行改编并同样获益? 利益相关者的观点和经验 —— 哪些群体支持哪个方案?	评价（例如，有效性研究，如随机对照试验） 评价（例如，观察性研究） 技术评估/成本效果评估 评价（例如，过程评估即评估一个方案是如何以及为什么发挥作用的） 定性研究（例如，使用访谈和焦点小组来了解什么对于公民来说重要）
3	促进和阻碍因素 —— 什么（和谁）会妨碍或帮助我们在恰当的人群中实现并获得预期的效果? 实施策略的益处、危害、成本效果等 —— 我们应该采用何种策略并在合适的人群中实现并获得预期的效果?	定性研究（例如，使用访谈和焦点小组来了解促进和阻碍因素） 行为/实施研究 见“选择方案”
4	所选择的方案是否能惠及那些能从中受益的人? 所选择的方案是否在足够的规模上实现了预期效果?	数据分析 评价

4.4 本土证据和全球证据的相互作用

决策者同时需要本土证据（例如，从他们自己的国家、州/省或城市学习到的东西）和全球证据（从世界各地学到的东西，包括不同群体和环境间的差异）。对于“本土”我们是指国家和次国家，其证据可表现为多种形式，包括本土数据分析、本土评估和本土实施研究。全球证据通常采取证据综合的形式，我们将在下文讨论此问题。

决策者可以从基于本土和全球证据的推荐意见中获益。从定义上来看，指南提供推荐意见。正如我们在介绍部分提到的，在危机时期，我们最初通常必须依赖于新出现的指南（例如，尽管我们还没有足够地了解危机，但在此期间需要好好洗手），然后依赖于更新的指南（例如，我们现在有证据表明带口罩可以减少病毒传播）。在任何时候，我们都需要对所谓的“逆转”持开放态度，也就是说随着证据的不断积累，原本认为有好处的方法实际上并不奏效，甚至会造成伤害。技术评估可以提供建议，也可以通过对可能影响本土决策的社会、伦理和法律因素的评估补充现有证据，从而提供一种证据支持。

建模是最常见的本土证据形式。然而，它可以提供一种在全球范围内综合最佳证据的方式，正如在气候行动、药品报销和宏观经济政策等高风险领域所做的那样。建模也可以为本土证据形式提供支持，其中建模者有效地充当了一种证据中介。这种情况就如同政府政策制定者利用许多特定管辖区的COVID-19模型来预测封城等方案可能产生的未来影响（以及最重要的不确定性）。如果做得好，这种应用或不应用证据综合效果估计的建模可以有助于系统地给出专家意见。

其他形式的分析，如政策、系统和政治分析，可以为本土和全球证据提供信息或加以补充，我们将在第5.4节讨论这些类型的分析。

优势	证据形式
本土（国家或次国家）证据 	 数据分析  建模  评价  行为/实施研究  定性见解
全球证据 	 证据综合
根据地方和全球证据得出的本土（国家或次国家）推荐意见或证据支持 	 技术评估  指南

全球证据



证据综合通过系统和透明的过程来识别、筛选、评估和综合解决同一问题的所有研究。其目的是为了全面了解现有的知识，包括其结果在不同人群（如女孩和年轻妇女）和环境（如低收入和中等收入国家）间的差异。对于方案选择的问题，现有的部分知识可以是在何种环境下何种方法对何种人有效。

与其他总结全球最佳证据的方法相比，证据综合方法有四个优势，比如由专家对科学文献进行非正式的叙述性综述：



通过确保纳入所有相关研究，并对高质量研究给予更大权重，从而减少被误导的可能



通过增加分析中研究参与者的数量，增加对预期结果的信心



通过呈现参与者和研究环境的信息，以及理想情况下研究结果如何因这些因素而不同，从而更容易评估全球证据在特定环境下的含义



通过确保每个人都能获取相同的“数据”，并清楚地报告这些数据是如何合成的，使得对现有证据提出质疑变得更加方便。

其中第一个优点有助于解决科学中有时被称为重复或可重复性危机的问题，即单个研究中的许多发现无法被复制或重现。这种危机在医学（例如，使用羟氯喹和伊维菌素治疗COVID-19）、经济学和心理学等许多领域均有所记录。更令人不安的是，不可重复的研究结果比可重复的研究结果被引用得更多，甚至在重复失败被发表后也是如此。(1)

基于模型的未来探索，以解决气候变化的“多维复杂度”社会问题，使用多种类型的证据，并借鉴强大的相互比较练习，为上述类型的证据综合提供了另一种范式。基于这两种最佳方法，可能建成一条卓有成效的前进道路。(2)

本土证据



本土（国家或次国家）证据可阐明是否存在本土问题及其原因，解决问题的方案是否具有本土可行性和可接受性，以及可能阻碍或有助于在适用人群中达到和实现预期效果的本土因素。对于决策者来说，“本土”的意义是多样的，即对某一个人来说，“本土”可能是他们的国家；对其他人来说，则可能是他们的近邻。我们将在[第4.5节](#)中讨论证据的本土适用性问题。

4.5 高质量证据和低质量证据的区别

不是所有证据都是高质量和可靠的决策依据。许多（但不是所有）形式的证据都有相应的工具，可用于帮助判断证据（单个研究或证据体）的可靠性。正如我们所述，这些工具通过评分或评级的方式帮助使用者了解对证据的确信程度。现在许多期刊要求作者遵循报告规范，如适用于随机对照试验的CONSORT和适用于证据综合的PRISMA。大部分期刊并未要求审稿人使用特定工具评价研究质量或推荐强度；因此发表在同行评审期刊上的并不一定都是高质量文章。

问题	回复
 研究（和指南）的质量（或可信度）各不相同	<ul style="list-style-type: none">已针对特定研究设计（如随机对照试验）、研究设计大类（如观察性研究、定性研究和证据综合）和指南开发了质量评价（或严格评价）工具——实例（RoB2、ROBINS-I、JBI清单、AMSTAR和AGREE II）见本章末尾附录（第4.16节）工具可以产生的结果包括总结性判断（例如，RoB2或ROBINS-I中的低偏倚风险），根据得分进行范围分组（例如，AMSTAR中的高质量），一组得分（例如，AGREE II的六个领域），或一组可以为总结性判断提供信息的考虑因素（例如，JBI清单）
 证据体的可靠度（或对证据体的确信程度）各不相同	<ul style="list-style-type: none">为解决同一问题（例如，干预措施对特定结局的影响或公民依附于特定现象的意义）的证据体开发了可靠性评价工具——见第4.16节的两个实例（GRADE和GRADE CERQual）工具可以得到一个关于可信度的总体判断，包括对真实效果与估计效果的相似程度（例如，GRADE中的高质量证据），或定性研究结果对研究内容代表程度的可信度（利用GRADE CERQual）对效果估计的可靠性进行总结性判断比证明干预措施“有效”或“无效”的统计学意义检验更有帮助（如果显著性水平为0.05，则20次试验中会有一次偶然情况发生）
 推荐强度各不相同	<ul style="list-style-type: none">已为指南推荐意见开发了推荐强度分级工具（例如，GRADE，除上述对证据体级别进行分级外）——见第4.16节的示例工具可以产生一个总结性判断，即大部分决策者是否会选择继续进行干预（例如，GRADE中的强推荐）或者大多数人是否需要仔细权衡干预的利弊
 证据的来源（或用于产生证据的方法）可能很难判断	<ul style="list-style-type: none">目前还没有广泛接受的工具用以评估以下内容的可信度：<ul style="list-style-type: none">专家，尽管在预测方面确实存在像“良好判断项目”这样的例子（我们将在本章后文中讨论专家意见；关于模型参数的专家意见，则在第4.16节讨论）用于生成某些形式证据的模型（我们在第4.7节讨论气候变化模型时，同时在第4.16节讨论这个问题）用于生成某些类型证据的人工智能算法，尽管已经出现了像TRIPOD这样的例子⁽³⁾

当证据被整合到数据平台、模型和其他格式中时，以及利益冲突起作用时，区分高质量和低质量的证据就格外具有挑战性。我们将在后面的[第4.12](#)、[4.14](#)和[4.16](#)讨论。虽然这不是本报告的重点，但区分高质量和低质量的“原始数据”也是一种挑战，比如UNICEP已经开发了数据质量框架来帮助解决这个问题（bit.ly/3DQQRRv）。

一些“一站式商店”，如社会系统证据（SSE）和COVID-19知证决策协作网（COVID-19 Evidence Network to support Decision-making, COVID-END）清单（在第4.6节中描述），使用了其中的一些工具，以便于决策者及其支持者可以关注高质量的证据综合，或者了解他们正在使用的是当前可用（如果不是高质量的）的最佳证据综合。

COVID-19大流行要求决策者在短时间内做出艰难的决策，最初的证据很少，且通常是间接证据。后来随着时间的推移，研究、证据体和推荐意见均通过一个稳健的过程生产或制订。为支持基于证据体（而非单个研究）的COVID-19决策，COVID-END在其“最佳”证据综合清单中列出了那些最新的（基于检索证据的日期）、高质量（基于AMSTAR工具）证据，并提供证据的可靠性评估结果（基于GRADE工具）。

并非所有证据均是高质量的，也并非所有全球证据均适用于特定情境。例如，证据综合只纳入在高收入国家开展的研究时，其可能在一些低收入国家中适用性有限。在基线条件、实际情况和制约因素以及本地系统的结构特征（例如，国家卫生系统或省/州教育系统）方面可能存在重大差异。SUPPORT工具可以同样帮助人们思考证据综合结果的本土适用性，并考虑即使发现不适用时，如何仍然得出见解。(4)

贝叶斯推理正在获得越来越多的关注，它可以重新绘制我们对于挑战和应对挑战的“心理地图”，不是用新的信息取代我们认为的所有知识，而是将我们的理解调整到一个适当的程度。这个程度取决于你对现有知识（某事为真的“先验”概率）的确信程度，以及你对新知识的确信程度。如果新知识来自于高质量的证据综合，其中包括了与你所处环境相似情境下开展的研究，则可以对新知识有更高的确信程度。



证据中介和生产者, *Gillian Leng*

是一位经验丰富的主管，领导技术评估和指南机构，支持政府、服务提供商和患者的健康和社会照护决策

英国多年来一直致力于鼓励证据综合和使用的有关工作，从第一个预防海员坏血病的随机对照试验，到最近新提出的促进证据在一系列政策领域使用的“What Works中心”。作为这次循证运动的一部分，在过去20年间，英国国家卫生与临床优化研究所（NICE）改变了证据在医疗实践以及更广泛的公共卫生举措和社会照护中的使用。

COVID-19大流行极大地增强了需要高质量证据为政策和实践提供信息的必要性，还突出了社交媒体和相关错误信息的负面后果。在这种情况下，全球证据委员会应对社会挑战的工作非常重要，应该被视为世界各地所有政策制定者的必备读物。

4.6 证据综合的覆盖面、质量和新近度

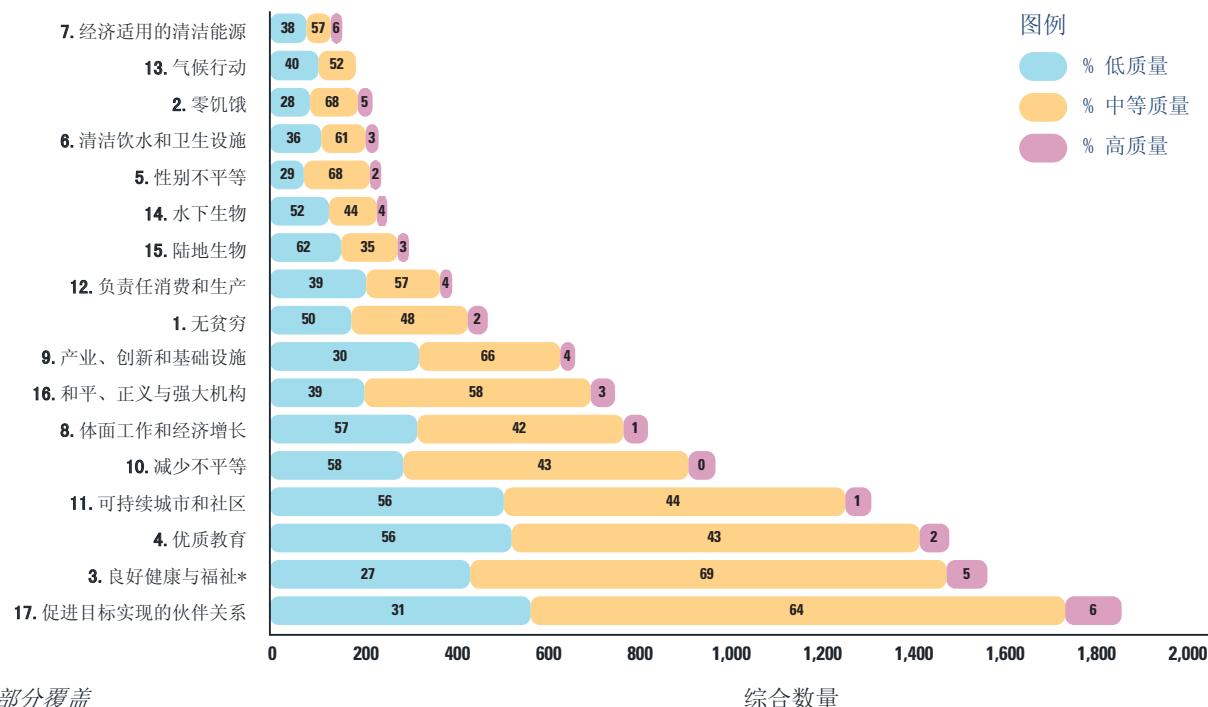
全球范围内的证据综合存在着对优先主题覆盖不全、（综合的）质量差异大、（对于检索本应纳入综合的潜在研究）新近度等问题。对两个“一站式商店”证据综合的分析说明了此问题的严重性。一个“站点”关注所有非健康的可持续发展目标，或称SDGs（社会系统证据），另一个则关注所有潜在的对COVID-19的响应（COVID-END最佳证据综合清单和该清单来源的大型数据库）。

SDG证据综合

截至2021年8月12日，在社会系统证据数据库中收录的4131份SDG证据综合中，包含了系统评价再评价、效果评价和解决其他问题的评价：

- 覆盖面不均衡，相对于每个SDG可提出的问题数量，其中七项SDG所涉及的证据综合数量相对较少（≤263）（目标2：零饥饿，目标5：性别不平等，目标6：清洁饮水和卫生设施，目标7：经济适用的清洁能源，目标13：气候行动，目标14：水下生物，以及目标15：陆地生物）
- 质量参差不齐，七项SDG的证据综合中至少有一半质量不高（目标6：清洁饮水和卫生设施，目标7：经济适用的清洁能源，目标9：产业、创新和基础设施，目标12：负责任消费和生产，目标13：气候行动，目标14：水下生物，以及目标15：陆地生物）
- 所有SDG的最后检索年份的中位数均为五或六年前（2016年或2017年）
- 大多数SDG的证据综合中，只有十分之一（12%）到五分之一（21%）包括至少一项来自中低收入国家的研究，其中一项SDG（目标9：产业、创新和基础设施）的证据综合包含中低收入国家研究的比例更低（3%）。

关注各个SDG的证据综合数量和质量在下面的柱状图中列出。



对于柱状图，请注意以下几点：

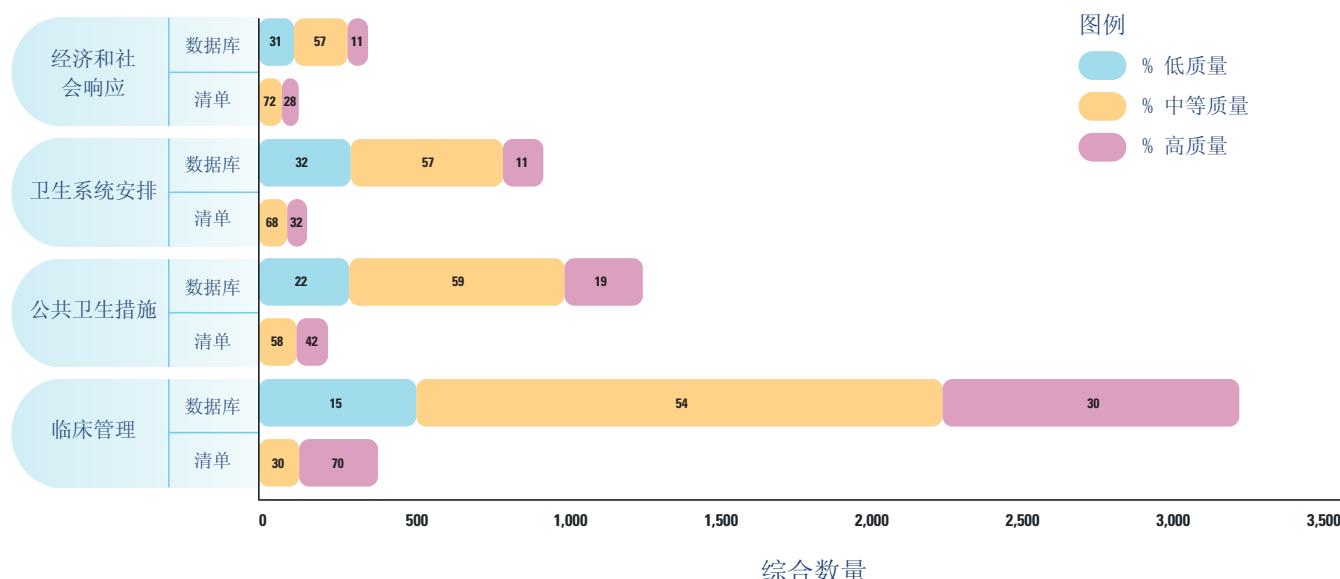
- 所有数字之和超过证据综合总数的原因是一个综述可能涉及一个以上的SDG
- 所涉及的证据综合数量：
 - SDG3数量明显不足，与健康有关的证据综合只有在同时涉及另一个SDG时才会被包括在内
 - SDG17数量显著超额，由于许多证据综合以另一个SDG作为主要问题，也将其他伙伴关系作为次要问题
 - SDG 7、13、14和15的数量可能被低估了，因为它们最近才被社会系统证据纳入关注
- 社会系统证据中，85%的证据综合已经完成质量评级。

COVID-19证据综合

截至2021年8月1日，在COVID-19全文数据库和562条COVID-END最佳证据综合清单中纳入了4256部COVID-19相关证据综合：

- 覆盖面不均衡：只有237部证据综合关注COVID-19的经济和社会响应（其中只有49部被纳入清单），而关注临床管理（3128部）、公共卫生措施（1148部）和卫生系统安排（818部）的证据综合数量要多得多
- 质量参差不齐：约四分之一（26%）的COVID-19证据综合为低质量证据，超过一半（56%）为中等质量证据
- 在COVID-19的四个响应类别中，三个响应类别的最后检索日期的中位数为世界卫生组织（World Health Organization, WHO）宣布大流行后的4.5个月内（2020年3月11日）。

关注临床管理的证据综合的中位检索日期更近——大流行宣布后12个月，分析完成前4.5个月（由于COVID-NMA动态证据平台上大量的药物治疗比较均为相同的检索日期）。关注各个COVID-19响应类别的证据综合数量和质量在下面的柱状图中列出。



对于柱状图，请注意以下几点：

- 所有数字之和超过证据综合总数的原因是一部综述可能涉及COVID-END分类类别中一个以上的类别
- 证据综合的质量等级需达到中等或高等，才能被考虑纳入COVID-END的“最佳证据综合”清单。

这些发现反映了评价（特别是随机对照试验）、证据综合和（评估和证据综合的）证据地图中的类似不足，可为决策提供参考：

- 教育：只有25%的试验受试者超过1000名（在1980–2016年期间开展的试验中，只有12%是在亚洲、非洲或中南美洲地区开展的）(5)
- 健康：只有16%的证据综合在其分析方法中纳入了质量评价（尽管70%的证据综合进行了质量评价），更普遍的是证据综合的报告质量差异很大 (6)
- 中低收入国家的可持续发展：在2010–2017年期间，有四个或更少的证据地图报告了与17个SDG目标中的八个相关结果，四分之一的证据地图没有以任何方式解决公平性的问题。(7)

其他此类评价工作则更为积极，例如，有一项研究指出社会工作中的740项随机对照试验表明这种评价方法在该领域确实是可行的。(8)

4.7 动态证据产品

决策者通常遇到的四种证据形式当前可以作为“动态”证据产品，这意味着它们会随着新数据的增加或新研究的发表而定期更新。许多这样的动态证据产品最开始是作为COVID-19证据响应的一部分。除卫生部门外，其他部门存在的数量较少。举例如下：

许多政府政策制定者和其他决策者已经开始期待COVID-19证据的定期更新，可能开始询问为什么不能为其他高优先级的社会挑战提供这样的产品。原因是这些挑战存在显著的不确定性，而且证据的产生很可能会解决这种不确定性。除其他创新产品外，人工智能的使用越来越多，这可能会使证据生产者在未来更容易满足这些更高的期望。然而，证据生产者将需要采取措施，以确保这些创新产品不会无意中延续或增加歧视的风险（例如，使用种族或与种族有关的变量，使某些群体处于不利地位）。他们还需要支持决策者恰当地解释和使用研究结果，特别是在进行因果推断时。

证据形式	动态证据产品示例
数据分析	<ul style="list-style-type: none">WHO的COVID-19面板提供了一套为应对COVID-19而采取的公共卫生措施其严格程度的数据分析，英国卫生安全局的监测报告 (bit.ly/3DeaS1c) 提供了一套关于英国COVID-19的数据分析，机遇与洞察力 (Opportunity Insights) 公司的经济追踪器提供了一套COVID-19对美国人民、企业和社区经济前景影响的数据分析经济合作与发展组织 (Organisation for Economic Co-operation and Development, OECD) 的经济活动每周跟踪报告提供了一套大多数OECD和G20国家经济活动的数据分析
建模	<ul style="list-style-type: none">欧洲COVID-19预测中心每周都会对每10万人中的病例和死亡人数进行预测——包括总体预测和国别预测，而卫生计量与评估研究所COVID-19预测中心每两周更新一次COVID-19的预测死亡人数模型，包括报告为因COVID-19死亡的人数和实际因COVID-19死亡的人数，可用于探讨特定国家的一系列情景（例如，关于口罩使用和疫苗接种）政府间气候变化专门委员会每五到七年提交一份评估报告，该报告借鉴了人类引起的气候变化模型、其影响和可能应对的方案，尽管严格来讲，这是一个通过稳健的模型间比较过程（每份评估报告由不同的科学家进行——见bit.ly/3wKQy8D的一个例子）的模型结果（可能是或不是动态的）的综合
证据综合	<ul style="list-style-type: none">COVID-END动态证据综合#6每两周针对COVID-19疫苗对抗变异株的有效性进行一次更新，COVID-NMA每周更新一次关于所有药物治疗COVID-19的证据综合（后来又增加了预防性治疗和疫苗）全球碳计划 (Global Carbon Project) 基于建模和实证研究，每年更新全球碳预算（气候变化下人为二氧化碳排放及其在大气、海洋和陆地生物圈中的重新分配）的五个主要组成部分的估计值及其相关不确定性
指南	<ul style="list-style-type: none">WHO关于COVID-19药物的动态指南每一至四个月进行一次关于COVID-19药物治疗的更新，澳大利亚国家COVID-19临床证据工作组每周为澳大利亚卫生专业人员更新COVID-19循证指南英国教育捐赠基金会为学校提供动态指南作为其教学工具包的一部分，例如涉及教学助理干预的指导

对COVID-END社区的列表服务讨论进行主题分析后，发现对以下问题存在不同观点：

- 对“动态”证据综合的理解（例如，“动态”状态的范围是否可以用一个时间范围而非“是”/“否”来进行更好地描述，是否应该为更新的频率设定一个下限）
- 什么时候应该开始更新，或者什么时候现有综合应该成为“动态”（例如，新的证据正在迅速出现，并且该证据可能解决决策者对某一高优先级主题的关键领域的不确定性）
- 什么时候可以停止更新（例如，证据不太可能改变我们已有的理解，而且该主题的优先排序被降低）
- 在哪里以及如何最好地传播（例如，期刊是否可以容纳一个过程，即最初经过同行评审的证据综合会定期更新，且不耽误其他的同行评审，决策者是否可以依赖在规定的时间内更新的承诺）

这些问题可能会成为未来几年激烈辩论的焦点。关于动态证据综合的理由和维护这些证据所涉及的问题，详见委员会一位委员参与共同撰写的简短说明。(9)

在第4.13节中，我们描述了作为COVID-19证据响应的一部分而持续推进的动态证据综合的一些关键特征。



证据生产者，Jan Minx

是一名以影响力为导向的学者，将创新证据综合方法引入气候变化和可持续性的国内政策建议和全球科学评估

我正在研究两种形式证据之间的接口：1) 证据综合，这是一种向过去取经的方法，被广泛用于卫生部门；2) 建模，旨在预测未来，并广泛用于气候变化领域。我强烈支持第19条建议，即我们需要向其他部门的证据小组学习。正如我们在该建议中指出的那样，Cochrane开创了许多方法，对探索什么对健康有益的研究进行综合。包括动态证据综合，IPCC也开创了很多方法来模拟人类在很长时间范围内引起的气候变化。Cochrane和IPCC可以相互学习，也可以向他人学习，其他人也可以向他们学习。

“

4.8 最佳证据与其他证据的对比（以及如何从其他证据中获得最大收益）

许多个人和团体提出了他们所谓的证据来应对社会挑战。特定国家（或次国家）环境下的“最佳证据”——来自于当前最佳研究的国家（或次国家）证据（例如，在该环境下学到的东西）和来自当前最佳证据综合的全球证据（例如，我们从世界各地学到了什么，包括它如何因人群和环境的不同而变化），需要区别于有时被作为证据的“其他证据”，如单个研究、专家意见、专家小组、研究兴趣小组、“包装为案例研究”的轶事、白皮书和辖区范围经验综合，这些其他的证据都会带来风险（下文第2列）。同时，有一些方法可以从它们那里获得更多的价值（下文第3列和第4列）。

我们在此不考虑那些通常作为研究证据的“其他证据”，如人们的生活经验（我们在第2.3节中结合共同设计的干预措施讨论）或原住民的认知方式（我们在第4.10节中对于原住民的这方面特征进行了更广泛的讨论）。

其他事物	存在的风险	解决方法	更好的改进措施
单个研究 (包括预印本) 	一手遮天*，或关注由作者及其媒体关系办公室或其他人积极推动的每项研究（如第3.7节中关于羟氯喹的高偏倚风险研究和现已撤稿的关于疫苗与自闭症之间联系的研究**）	要求使用得到广泛认可的质量评价工具对研究进行批判性评估（以了解偏倚风险），并需要认识到具有统计学意义的显著性水平（0.05），在20次试验中仍旧会有一次偶然情况发生	将该研究添加到“动态”证据综合中，使其能够与解决相同问题的其他研究一起被了解（或将其与最佳全球证据放在一起，视为许多国家或次国家证据之一）
专家意见 	“巧妇难为无米之炊” / “基于权威人士”（而不是以证据为基础）的决策，或关注那些凭借坚持、声誉或其他因素获得最大关注的人（就像被广泛观看的电视节目《少年监狱之旅（美版变形记）》（Scared Straight）中预防犯罪的节目一样，即使在证据综合***发现有害无益的证据之后，仍旧可能发生类似于电视节目中的情况）	请专家分享其观点所基于的证据（最好是证据综合），以及用于识别检索、评价、筛选和综合证据的方法	让专家参与研究特定的证据综合对特定辖区的意义，或用不同形式的证据挑战思维方式****（或问专家什么证据能让他们相信自己是错的）
专家小组 	GOBSATT (good old boys sitting around the table) 提供他们的个人意见	要求小组成员分享他们的意见和推荐意见所依据的证据（最好是证据综合），以及用于检索、评价、筛选和综合证据的方法	在小组（或秘书处）中增加方法学专家，预先传阅本土（国家或次国家）和全球的最佳证据，提供强有力的审查，并明确哪些推荐意见是基于何种强度的证据

* 我们用“一手遮天”（即狗撵汽车（不懂科学））比喻分享和评论每一项引人注目的新研究。

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

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

**** 这种挑战在军队中被称为“红队”。

辖区范围经验综合 	<p>“群体思维”，或许多管辖区域的人依靠一个管辖区域的人，他们愿意分享他们的经验和创新，但尚未对其进行评估</p>	<p>要求或寻找任何可用的支持性证据或产生证据的计划</p>	
研究兴趣小组***** 	<p>研究人员根据个人价值观和偏好或其专业兴趣采取行动</p>	<p>询问小组为什么他们的价值观和偏好比我们所服务的公民更有价值</p>	<p>鼓励他们基于高质量证据综合提出要求</p>
“案例研究” 	<p>轶事经验的名称意味着有严格的方法支持</p>	<p>询问作者选择案例的标准，使用的数据收集混合方法，以及为保证严谨性使用的分析方法和其他方法</p>	
白皮书 	<p>从表面上看，指在得出政策倾向的声明时使用了证据的隐含或明确的断言</p>	<p>要求政府的领导或顾问分享他们投入资源和建议所基于的证据，以及用于检索、评价、筛选和综合证据的方法</p>	

*****注意，社会利益集团也可能基于其价值观和偏好倡导行动时援引证据，在这种情况下，第4栏中的回答可能是恰当的。

4.9 形成如何看待证据的环境

历史、社会和文化环境均会影响人们对证据的看法，例如，种族化社区（我们在第1.7节介绍的PROGRESS-Plus中的R）和妇女（PROGRESS-Plus中的G）等。一些环境与过去产生证据的努力直接相关，而另一些则与过去将特定群体描述为“异类”的行为有关，这可能表现为这些群体对任何声称支持他们或关于他们的证据持怀疑态度。如果我们要以采取行动的方式生产和传播证据，就必须了解这些环境。（10; 11）

正如我们在第4.10节中所提到的，环境以及原住民独特的权利和认知方式，也可以塑造其对证据的看法。环境也可以决定错误信息是否泛滥以及如何泛滥，这也是第4.11节的重点。

环境示例	证据生产与传播的潜在影响
------	--------------

与过去美国生产证据的努力直接相关

-  对患有梅毒的黑人男子不予有效治疗，以便监测未经治疗的梅毒的发展情况 (bit.ly/3DeaH9x)
-  心脏病治疗试验不纳入妇女，但其结果却被认为适用于妇女 (bit.ly/3olxgTH)
-  对学生进行标准化测试的方法不利于有色人种学生，特别是那些来自低收入家庭的学生 (bit.ly/3wDICGk)

更加关注研究（或没有研究）什么，由谁来研究（例如，由来自不同环境的人组成的研究小组），如何研究（例如，以伦理为基础、以公平为导向的更多参与方法），以及为什么要研究（例如，找出可以利用的优势）

与过去新接纳特定群体的国家将其描述为“异类”的行为有关

-  将中国移民虚假描绘为肮脏的和有病的，以此作为在旧金山社区特别严格执行卫生条例的理由 (bit.ly/3qzeJFV)
-  关于撒切尔时代英国黑人成为国家问题“外部”根源的隐含信息出现在书籍和电影中，并被一些观众作为事实接受 (bit.ly/3naBa2n)
-  媒体报道将某些人群，如欧洲的穆斯林移民和美国入侵伊拉克后的伊拉克被拘留者，视为已经“丧失”（失业、饥饿和监狱）的人群，不值得社会保护 (bit.ly/3wGrKyE)

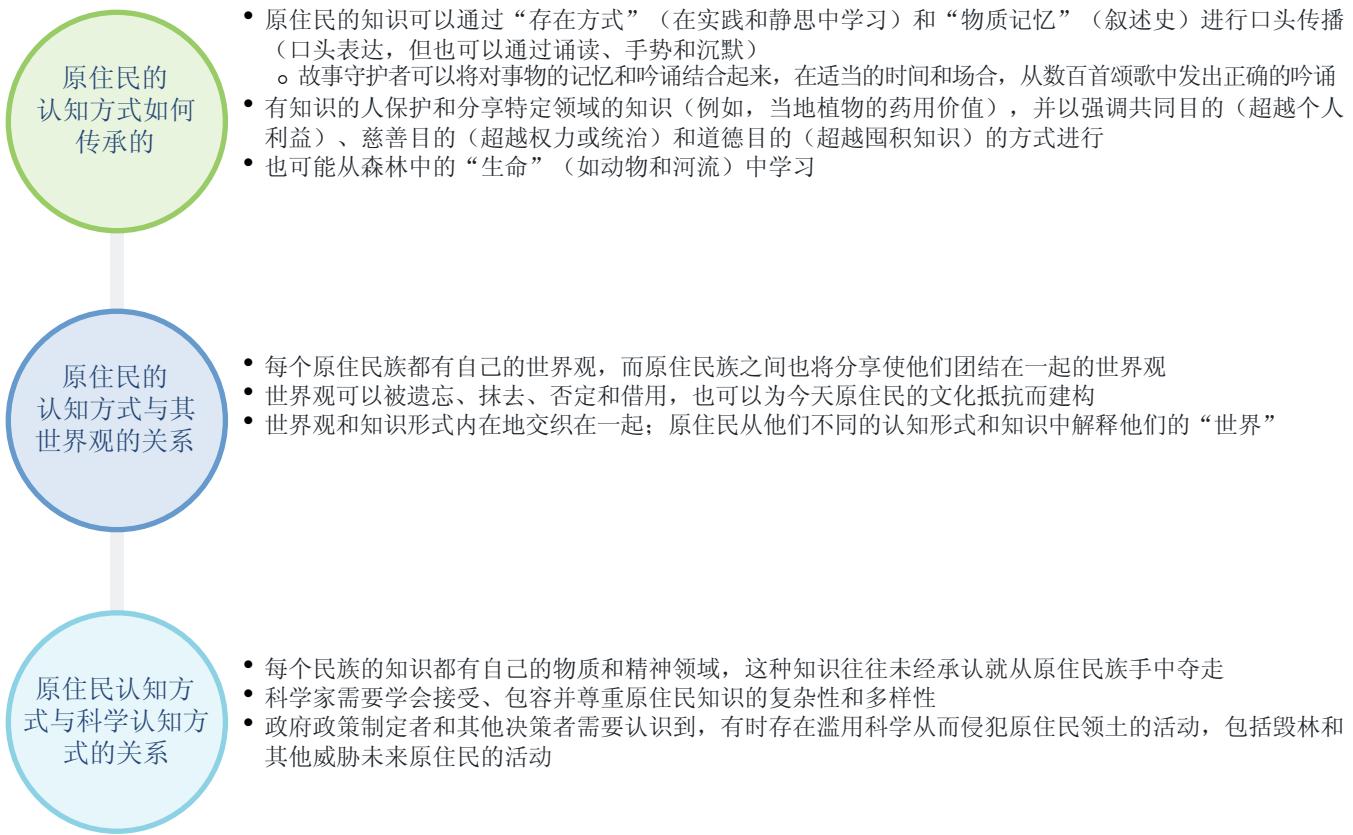
给予各种媒体对证据描述的更多关注，并利用这些见解，设法预测各群体对针对他们或关于他们证据的反应，或了解他们为什么会有如此反应

4.10 原住民权益及其认知方式

作为承认和确保原住民权益更广泛转变的一部分，许多政府政策制定者、研究人员和其他人开始接受，原住民应该控制数据收集过程，并享有和掌握这些证据的使用方式。基于原住民数据所有权（ownership）、控制权（control）、访问权（access）和占有权（possession）原则（也称OCAP原则），国际原住民数据主权利益集团制定了《原住民数据管理的CARE原则》（CARE Principles for Indigenous Data Governance, C: 集体利益（Collective benefit）、A: 掌握权（Authority to control）、R: 责任（Responsibility），E: 道德（Ethics））。这些原则旨在补充管理科学数据的FAIR指导原则（FAIR: 可查找（findable）、可访问（accessible）、可互操作（interoperable）和可重复（reusable））。目标是让原住民数据的管理者和用户做到“公平（FAIR）”和“关怀（CARE）”。这种证据相关的权利应被理解为通过《联合国原住民权益宣言》（United Nations Declaration on the Rights of Indigenous Peoples）确立的一系列更广泛权利的一部分。

原住民的认知方式是一个反映原住民学习和教学方法多样性和复杂性的术语。这种多样性源于许多已形成自己认知方式的原住民或民族，这些方式在他们被殖民化之前的几个世纪以及此后的时间里不断发展。复杂性由许多因素造成，包括知识来源众多。虽然原住民的认知形式存在共性（例如，将个人视为与周围的人和土地相互联系的整体观点），但最好不要一概而论。下表为讨论原住民认知方式的切入点，由专员Daniel Iberê Alves da Silva（M'byá Guarani人）指导编制，其传记见附录8.2。正如我们所述，进一步的讨论应始终由原住民主导。

领域	细节
原住民认知方式的来源	<ul style="list-style-type: none">知识来自个体与世界的关系，既有物质层面，也有不可分割的精神层面知识的来源包括植物、动物、其他人类和土地元素（如山脉和河流），以及理想、灵魂和精神世界的其他表现形式。例如，水的世界包括湖泊和河流及其蕴含的精神。更笼统地说，在原住民文化诞生和发展了几个世纪的物质领土上，存在许多因精神赋予他们“生命”（这使得强迫移居特别具有破坏性）的“物质”物质环境可以作为精神层面的提示或灵感，帮助形成行为模式（例如，观察河流的流动可以让观察者找到解决某个问题的方法）与掌握如何完成的“秘密”的人一起学习
原住民认知方式的特点	<ul style="list-style-type: none">原住民的知识是整体性的，与每个民族的历史、文化和领土有关（例如，他们的创世神话及其与其他“生命”的联系）“知识”体现在个人的经验或“存在”中（例如，成人仪式（人生重要转折）是个体“体验”和发现事物本质的必经之路）知识在原住民族内部、民族与民族之间以及与他人共享，并随着时间的推移不断完善（例如，今天独木舟的制作方式已与两个世纪前不同）知识可以通过自己的感受（传统意义上指身体感官，但也可以通过衣着、饮食、绘画和音乐）、语言（可以说的）和沉思（不能说的）来获得
在原住民的认知方式中“事物”如何分类	<ul style="list-style-type: none">不同个人和原住民族因其文化、历史或领土对分类的看法并不相同（例如，同一种植物，一个原住民族根据其治疗用途进行分类，而另一个原住民族则根据其与死亡的联系进行分类）分类可以随着时间的推移而改变（例如，一些植物曾被拟人化），并且根据其内在的“精神”来理解



4.11 错误信息和信息疫情

错误信息是指不考虑误导意图而传播的不正确信息。虚假信息是指故意传播的错误信息。例如，政敌或外国政府可能会为了实现某一特定目的而参与传播虚假信息，如选举优势或破坏公众对民主制度、独立媒体和科学知识的信任。有组织的团体可能出于其他目的，如赚钱或推进某种意识形态。由于很难证明传播的意图，所以我们在本文使用“错误信息”一词。虽然错误信息已经伴随我们几个世纪了，但互联网已经改变了它的规模、驱动因素、后果以及可能的应对措施。

在新型冠状病毒肺炎（COVID-19）大流行期间，人们开始使用“信息疫情（infodemic）”（或“错误信息疫情（mis-infodemic）”）一词来捕捉病毒的快速传播和COVID-19相关及其预防、管理和减轻经济和社会影响措施的错误信息快速传播之间的相似之处。现有与疫苗有关的错误信息往往在COVID-19疫苗上市后被重新控制，并发起了许多新的反疫苗接种活动。

2020年，由国际电信联盟（International Telecommunication Union, ITU）和联合国教科文组织（United Nations Educational, Scientific and Cultural Organization, UNESCO）资助，可持续发展宽带委员会（Broadband Commission for Sustainable Development）发布了一份在尊重言论自由的同时打击数字错误信息的报告。(12)

该报告描述了错误信息生命周期的五个阶段：



煽动者和受益者，出现了关于动机（和上述目标）的问题



代理人，出现了关于技术的问题，如机器人和伪造帐户或虚假身份



信息，涉及形式的问题，其中三个常见的形式是：

- 情绪化的主张和叙述，其通常混合了情绪化的语言、谎言或不完整信息、个人观点和真相要素
- 伪造、去语境化或欺诈性修改的图像和视频，以及合成音频
- 虚构的网站和已污染的数据集



中介机构，出现了关于平台（如暗网、社交媒体、消息和新闻媒体）和正在被利用的平台功能（如算法和商业模式）的问题



目标人群和解释者，出现了关于谁受到影响（例如，公民、科学家和记者等个人；研究中心和新闻机构等组织；黑人社区和原住民族等社区；以及选举流程等系统）以及他们如何应对（例如，通过忽略或分享揭穿错误信息）的问题

该报告将错误信息与模仿和讽刺区分开来，模仿和讽刺既可以误导那些没有能力识别错误信息的人，也可以通过强调错误信息的荒谬成分进行反击。

可持续发展宽带委员会的报告也提出了对错误信息的潜在应对方法，并提出了与言论自由权利交叉的例子。联合国教科文组织的报告指出了这些应对措施的潜在互补性，并指出有必要确保所用任何应对措施的一致性。

监测和事实核查

- 包括对错误信息的监测和揭露（例如，被推翻的主张）以及对新的主张进行事实核查
- 由独立组织雇用训练有素的专业人士进行判断（即使有自动化的帮助）可减少侵犯言论自由权利的风险

可靠度标识

- 包括内容验证工具、网络内容指标、标示（指向可靠的证据来源）和网站可信度标签

教育性

- 包括培养公民的媒体和信息素养（例如，批判性思维和数字验证技能）以及记者的信息素养

策划性

- 将用户指向官方可信的证据来源，供新闻媒体、社交媒体、讯息通信和搜索平台使用
- 可能被误用为一种私人审查形式

技术和算法

- 涵盖可以识别错误信息、提供附加环境并限制传播的人类学习、机器学习以及其他人工智能方法
- 上诉过程的自动化可能会侵犯言论自由权

反错误信息运动

- 包括专门机构制订挑战错误信息的反击言论，并动员线上社区传播高质量证据

规范性

- 通常由政治和社会领导人公开谴责错误信息行为，并提出相应解决建议

经济性

- 包括广告禁令，特定内容（例如COVID-19内容）去货币化和其他消除促使错误信息产生的方法

立法和其他政策

- 包括将错误信息行为定为犯罪，指导互联网通信公司删除内容，以及为可靠的信息来源提供物质支持
- 可能会被滥用于削弱合法的新闻报道并侵犯言论自由权

调查性

(可为立法和其他应对措施提供信息)

- 检查煽动者、传播程度和手段、涉及的资金以及受影响的社区

尽管存在许多这样的证据综合，但该报告并未涉及支持这些措施的证据。例如，一个较早的中等质量证据综合（AMSTAR评分7/11，检索时间为2017年）发现：纠正错误信息（即措施1）对错误信息的相信有一定影响（对健康的影响比对营销或政治的影响更大）；反驳比预警更有效；呼吁一致性比事实核查和呼吁可信度更有效。(13) 此处目的并不是提供关于这些措施的知识现状，也不是探讨可能支持这些措施的错误信息心理学，而是表明已有错误信息应对措施的证据综合，同时需要动态证据综合。动态证据综合可以提供对已知情况不断发展的理解，包括如何因群体（例如，更容易受到错误信息影响或持有特殊信仰体系的群体）和情境（例如，两极分化的社会）的不同而变化。

正如我们在背景中所讨论的那样，如果我们能够继续为使用证据培养能力、机遇和动力（在这种情况下是为解决关于社会挑战的错误信息），同时运用判断力、谦逊和同理心，那么其结合将很好地为我们服务。即使我们依靠卫生部门常用的严格评价和可靠的自我校正系统，我们也可以做得更好。正如Ross Douthat在其莱姆病生活回忆录中所观察到的，我们需要更多具有以下两种世界观的人和机构：1) “接受现代科学的核心成就，至少对民粹主义信息源和权威机构源持相同的怀疑态度，并拒绝盲从”；2) “认识到我们的机构在各个方面的失败，有更广泛的、符合当前学术官僚路线的经验……”(14) 我们大多数人都在医学等领域受益匪浅，这些领域结合了严格评价和相当可靠的自我校正系统。但像Ross Douthat这样的人却没有，他指出：“与七年前相比，我对宇宙的态度更加开放，对任何声称共识的东西都更加怀疑。但我试图不让这种开放和怀疑的混合演变成偏执的局外人群体思维。”(14)

4.12 卫生研究系统的局限性

在COVID-19大流行之前，一组研究人员记录了卫生研究系统的局限性。为更好地满足决策者的需求，他们呼吁重组该体系，包括支撑该体系的结构（如Cochrane等全球合作机构）和激励措施（如来自大学、资助者和期刊的激励）。(15–17) 他们主要关注决策者常遇到的三种证据形式，即原始研究（特别是评估类研究，尤其是随机对照试验）、证据综合和指南（以及较少的卫生技术评估）。

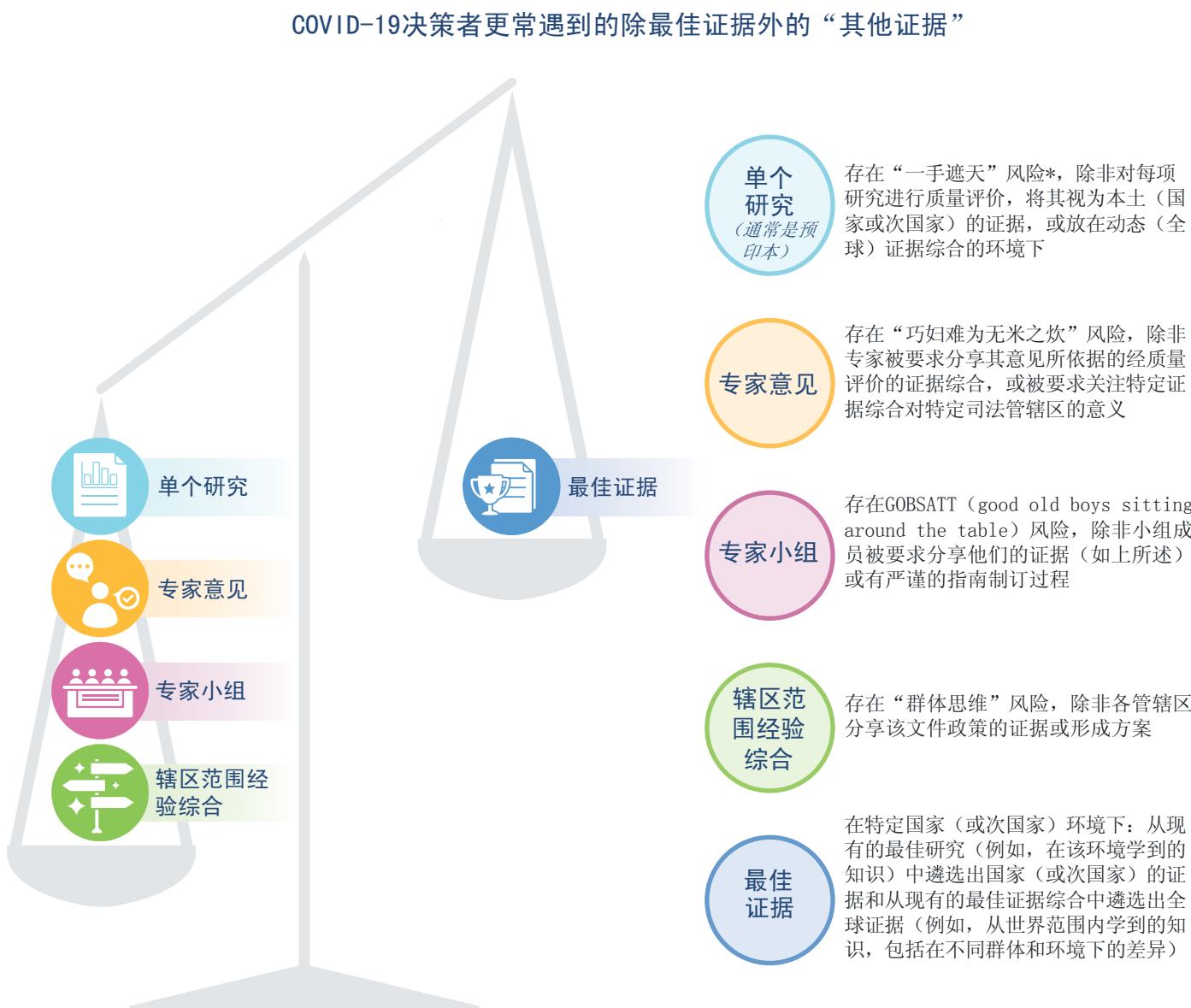
虽然一些局限性在COVID-19证据响应中变得更为明显，但在应对大流行中也产生了解决许多局限性的典型案例。尽管研究人员最初关注的是卫生挑战和特定的证据形式，但许多见解也同样适用于其他社会挑战和其他形式的证据。换言之，对于与本节所述完全不同的社会挑战和证据形式，也需要开展类似的工作。例如，政府间气候变化专门委员会（Intergovernmental Panel on Climate Change, IPCC）在其关注领域的全球协调方面发挥了巨大作用，并推动了长期建模的新方法。然而，IPCC也可以获益于通过对气候变化应对方案的事后评估来完善这些方法。

新冠肺炎大流行前卫生研究系统的局限性	在COVID-19证据响应中突显局限性的示例	在COVID-19证据响应中解决局限性的示例
缺乏证据社区的全球协作，每个社区都使用系统和透明的方法以及一系列完整的数据来源（例如，研究注册网站、监管机构和官方机构数据库）理想地应对全球优先挑战	<ul style="list-style-type: none">COVID-END全球审查工作组优先考虑的许多主题从未通过一部或多部“最佳”证据综合得到解决低信噪比：根据解决具体决策相关问题、证据检索的时间、证据综合的质量和GRADE证据分级可用性这四个标准，COVID-END（截至2021年11月7日）清单中关于COVID-19的近11000部证据综合大约仅有600部“最佳”证据综合	<ul style="list-style-type: none">为努力减少重复并加强协作，COVID-END与55部领先的证据综合、指南制订和技术评估团队以及公民合作伙伴和证据中介机构进行了合作PROSPERO鼓励拟注册COVID-19证据综合计划书的研究人员检索已有的计划书，如果重复，则选择新的主题（尽管138个团队仍在推进已由其他57个团队之一注册的主题，包括14个羟氯喹主题和7个托珠单抗主题）GloPID-R（全球传染病防治研究合作组织）协调主要研究资助组织快速资助COVID-19原始研究
证据社区对维持动态证据综合的关注不足。动态证据综合针对应对优先挑战的所有干预措施（例如，网状Meta分析，而不仅仅是单纯的配对比较）进行了审查	<ul style="list-style-type: none">只有13%的COVID-19证据综合自我定义为是动态证据综合（而COVID-END清单中，52%“动态”证据综合作为“最佳”证据综合标准），三分之二以上的证据综合涉及临床管理（而不是公共卫生措施、卫生系统安排以及经济和社会对策）只有21%的COVID-19动态证据综合进行过一次更新（在第一次发表后），8%有2次更新，13%有2次或更多的更新，而第一次检索与更新检索间隔的平均和中位时间分别为49天和31天许多COVID-19证据综合为解决单一药物治疗的问题，因此COVID-END清单转变为主要依赖COVID-NMA和其他比较所有药物治疗的研究（仅包括可获得所有预后因素的预后研究）	<ul style="list-style-type: none">四个证据社区维护了所有高质量的药物治疗动态Meta分析，其中一个（COVID-NMA）支持每周更新偏倚风险评价和GRADE证据分级结果

<p>证据社区对识别干预措施危害和益处的关注不足 (更普遍地说，包括更广泛的研究设计和数据类型)</p>	<ul style="list-style-type: none"> 当前研究和综合使人们很难理解如何看待关于某些疫苗接种者发生血栓的报告 	<ul style="list-style-type: none"> 一个COVID-END团队进行的一项系统评价完成了与接种疫苗相关的暂时性血小板减少症因果关系评估
<p>缺乏对个体参与者数据的共享，也没有利用这些数据检查调查结果如何因受试者类型、环境或其他因素的不同而变化，以及如何更好地进行个性化或情境化干预</p>	<ul style="list-style-type: none"> 许多报告记录了缺乏个体参与者数据的共享（例如，疫情初期的一项对140项研究的综合发现，仅有的一项研究的数据进行了共享——见 bit.ly/31WQUxM） 	<ul style="list-style-type: none"> COVID-19知识加速器改进了跨平台共享可计算的证据表达和指南推荐意见所需的方法，Vivli 扩展了其平台以共享COVID-19试验数据
<p>证据社区中未纳入来自所有相关证据团体（例如，进行原始研究的研究人员，如试验人员、证据综合者和指南制订者）、所有相关类型的决策者和所有相关类型的证据中介机构代表</p>	<ul style="list-style-type: none"> 许多报告描述了在COVID-19研究中公民的参与度低于疫情前的其他研究，同时在疫情初期没有提供证据综合的简明摘要（例如 bit.ly/3kwCHhr） 	<ul style="list-style-type: none"> 国家COVID-19临床证据工作组邀请了很多医疗专业人员（及其协会）和患者参与到动态指南中，并与主要维护动态网状Meta分析的证据团队合作 许多小组参与建模，以帮助决策者在基于现有证据和专家意见的可用方案（如封城）以及环境提供的某些条件下进行选择 许多团队应决策者的要求（在许多COVID-END快速综合的情况下，与公民合作）撰写了符合实际情况的快速综合
<p>证据社区缺乏使用一系列的新方法来提高其工作效率和及时性（例如，机器学习和众包对其工作的贡献）</p>	<ul style="list-style-type: none"> 截至2021年7月，已有超过18000项研究被上传到一个预印本平台（medRxiv），这极大地缩短了出版时间（与此同时，由于缺乏同行评审而存在不确定的危害） 在纳入183份报告的中等质量概况性评价（bit.ly/3D7bTeV）中发现许多在COVID-19响应中使用机器学习的例子，但在疫情初期并没有被广泛使用 	<ul style="list-style-type: none"> L*VE（动态证据综合再评价）使用机器学习维护原始研究和证据综合的数据库，EPPI中心使用机器学习来维护动态证据地图
<p>‘原始研究的空白、质量和透明度的报告不足（包括利益冲突）是反馈循环的一部分，旨在支持学习和改进——详见框1：(17)</p>	<ul style="list-style-type: none"> 许多原始研究的结果都是通过媒体发表的，而不是通过可以严格评价的完整研究报告 许多报告指出原始研究存在中高偏倚风险（例如，713篇中有81%的文章从10516篇COVID-19文章池中纳入原始的患者数据（bit.ly/3HiI90X），并因学术不端而被撤稿 COVID-END编制的报告针对证据综合缺乏时效性（在全文数据库和“最佳”证据综合数据库中，分别有91%和61%是180多天前完成的检索）、中低质量（分别为75%和55%）以及缺乏证据概况（分别为81%和42%）的报告，以及快速综合比完整综合更有可能是低质量的（43%对比13%） 	<ul style="list-style-type: none"> RECOVERY（recoverytrial.net）和世界卫生组织COVID-19团结治疗试验为超快速、高质量、多国家试验的COVID-19药物治疗提供了平台 COVID-19证据警示概述了基于质量评价的原始研究

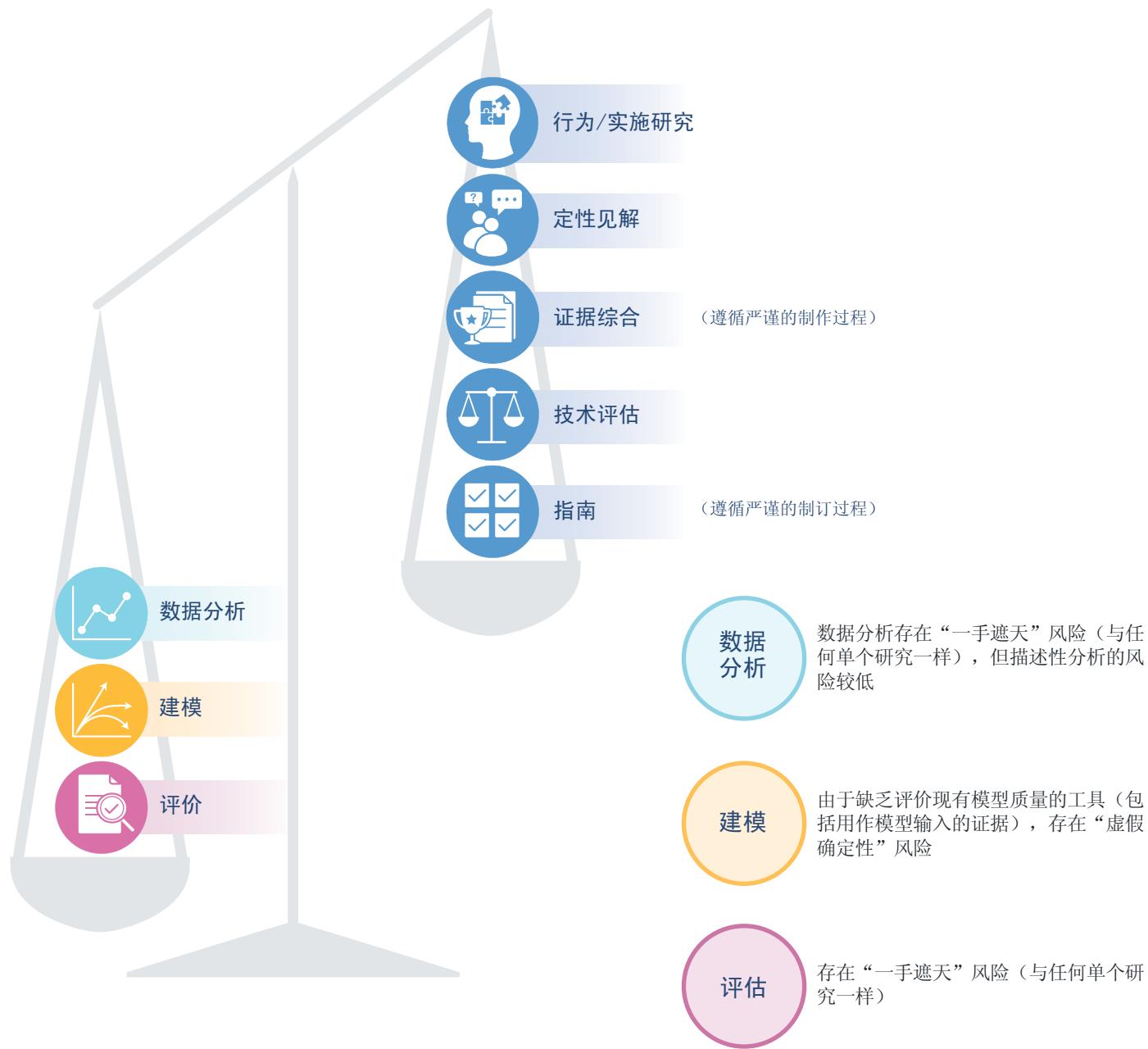
4.13 众多COVID-19证据支持系统的不足

COVID-19大流行是一场全球性危机，需要政府高层针对几波“浪潮”快速做出决策，同时还面临着证据的显著不精确性和快速演变（通常是间接的）。在许多地区，COVID-19大流行期间证据似乎在政府的政策制定中发挥着比过去几十年更显著的作用。尽管如此，错误信息盛行，公民和其他利益相关者难以理解为何证据会随着时间的推移而改变。相比最佳证据，“其他证据”往往更受关注，一些形式的证据往往比其他形式的证据更受关注。我们在[第4.11节](#)中讨论了错误信息，并为[第4.8节](#)（最佳证据以外的“其他证据”）、[第4.2节](#)（证据形式）和[第4.5节](#)（高质量证据和低质量证据的区别）中使用的术语提供了更多的环境信息。



*如[第4.8节](#)所述，我们用“一手遮天”（例如，即狗撵汽车（不懂科学））来比喻分享和评论每一项引起人们注意力的新研究。

COVID-19决策者更常遇到的证据形式



任何管辖区的领导者均可以采用《证据委员会报告》（Evidence Commission report）来将COVID-19证据响应较好的方面系统化并扩大卫生领域以外，也可以解决许多未解决部分。作为系统化证据响应较好的一部分，这些领导者需要从COVID-19时代关注速度和尽可能高质量的证据（“足够快且高质量”）过渡到速度平衡、质量（例如，等待即将生产的证据）和可持续性上（例如，保持正常工作时间且不搁置其他工作）。

4.14 国家证据基础设施的理想特征

每个国家都有一个国家证据基础设施，包括许多与证据相关的结构和流程。在国家证据基础设施中，我们区分了证据支持系统、证据实施系统和研究系统。加大对证据支持系统的重视并对证据实施系统持续保持关注，是未来使用证据应对社会挑战的关键。



决策者是使用证据的人，而研究人员则是开展研究的人。当决策者，尤其是政府政策制定者和组织领导者提出问题时，需要使用现有的证据及时进行支持。决策者，特别是专业人士和公民需要强有力的证据支持，以实施所需的改革。同时，研究人员需要发明新的产品和服务，开发新的思维方式并批判现状。还需鼓励研究人员更积极地与决策者接触，以确保相关性和适用性，更有效地利用技术使研究过程更有效率、更透明地报告研究结果，而不是“杜撰”，并创建生产证据的模式，便于决策者获取、理解并采取行动。从研究中出现的“即将进入全盛时期”的证据，可以被纳入证据支持和证据实施系统。

证据支持系统	驱动因素	辅助因素
基于对国家（或次国家）环境（包括时间限制）的理解，以需求为导向，并侧重于以公平敏感的方式为决策提供符合情况的证据 基础设施示例： <ul style="list-style-type: none">• 证据支持协调办公室（适用于所有政府部门，在关键部门或部委设或不设额外办公室）• 证据单位，在八种证据形式中各有专长（例如，行为见解单位）• 提出证据需求并确定其优先级，在规定的时间内找到并打包满足这些需求的证据（并构建额外证据作为持续评估的一部分），培训使用证据的能力（例如，证据使用研讨会和手册），促进证据使用（例如，内阁提交清单），并记录证据使用（例如，证据使用指标） <p>虽然此类基础设施与政府政策制定者和大型组织领导人最为相关，但类似的基础设施也可以为小型组织领导人以及专业人员和公民量身制定</p>	受用于 <ul style="list-style-type: none">• 本土证据中介机构• 来自Cochrane、Campbell和其他机构与证据相关的全球公共产品（例如，全球标准和证据综合的开源出版物）• 来自联合国和其他多边组织的技术援助，包括其国家、区域和全球办事处	辅之以： <ul style="list-style-type: none">• 预测未来证据需求的前瞻性倡议• 创新中心，以发明新产品和服务，对其进行评估并扩大可以通过市场或公共采购增加价值的产品和服务规模

	驱动因素	辅助因素
证据实施系统*	与上述内容类似	政府政策制定者和组织领导人使用现有杠杆以支持实施（例如，将推荐的产品和服务添加到福利包中，并规定对反映遵守建议行动的指标公开报告）
研究系统	与研究相关的全球公共产品（例如，开放科学倡议）	辅之以：政府政策制定者和组织领导人使用现有杠杆奖励某些活动（例如，英国卓越研究框架（Research Excellence Framework）的机构评估活动）

*我们使用“证据实施系统”一词与证据支持系统进行区分。近期，证据实施系统被称之为证据生态系统。(18) 我们之所以避免使用此术语，是因为它既混淆了那些习惯于生态系统字面含义的人，也因为它没有抓住这个系统对实施的关注。如果我们使用“证据生态系统”一词，我们可能会将其应用于证据支持系统和证据实施系统的组合。

在上述第一行的基础上，理想情况下，证据支持系统应具备以下特征：

- 最佳证据，并以下述方式支持政府政策制定者以及组织领导人、专业人士和公民的决策：
 - 充分了解环境，包括在何处以何种方式进行决策，决策的时间限制及决定是否向有需求的人提供正确的产品和服务的已有制度安排，以及在决策中使用证据的能力、机遇和动机
 - 响应与决策相关的需求、时间限制以及对产品和流程形式的偏好
 - 反映将最佳证据与所提问题相匹配的承诺，并致力于研究证据对特定决策的意义（例如，将证据置于环境中），包括因人群和环境的不同而变化（例如，将公平的视角引入证据及其被如何看待）
 - 实现判断力、谦逊和同理心，并适当关注利益冲突的识别和管理
 - 通过政府内部人员以及与政府以外的证据中介机构和生产者（例如，国内证据中介、全球公共产品和技术援助的提供者）建立战略伙伴关系，并以系统和透明的方式实现
 - 辅之以在联合国称之为“变革五重奏”的两个部分（即战略远见和创新）。(20)
- 变革五重奏的另外三个部分为数据分析、行为/实施研究和评估（“绩效和结果导向”），它们在八种证据形式 中均已体现。

一些政府已选择通过立法等方式正式确定证据支持系统的各个方面。美国国会成立的循证决策委员会（Commission on Evidence-based Policymaking, CEP）(21) 提出的建议为《证据法》（Evidence Act）提供了证据。总统和国会预算办公室的后续备忘录有助于支持该法案的实施。这些努力与证据委员会一样，均关注所有类型的社会挑战，但不同的是，它们仅关注一类决策者（即政策制定者，这里指美国联邦政府），并仅关注两种形式的证据（数据分析和评价）以及生产新证据而不是更好地利用现有证据（例如，通过证据综合）。联合国系统的一些部门选择通过有关加强证据支持系统的决议。在东地中海区域，WHO区域委员会也通过了一项针对卫生部门的类似决议。(22)

4.15 以证据形式区分的全球委员会报告

自2016年1月以来，全球委员会发布了70份报告，其中仅一份报告在描述其委员时，从决策者常遇到的八种形式证据中特别遴选了专业知识。

在委员会报告的方法部分明确报告他们在自己的工作中借鉴了这些形式中的任何一种证据时，建模是最常见的证据形式（13份报告），其次为证据综合（6部）和技术评估/成本效果分析（5份）。作为对方法部分的补充，对参考文献的分析发现：

- 70份报告中64份有参考文献清单
- 在此64份报告中，只有32份报告至少引用了一部证据综合
- 仅有3%（526/17605）的引文可以根据其标题判断为是证据综合
- 每份报告的证据综合引用频次的平均值和中位数分别为8.2次和1次。

我们还分析了《2019年全球可持续发展报告》（Global Sustainable Development Report 2019）的引文清单，该报告由联合国秘书长任命的独立科学家小组撰

写，因此，人们可能会认为这份报告将优于其他。（20）然而，在这份报告中，仅有1.8%（17/941）的引文可根据其标题判断为是证据综合。当引用证据综合时，并不清楚在选择使用它们时是否考虑了检索的质量和新近度。例如，引用的证据综合中有三部涉及卫生工作人员招聘和留用这一特定主题，但通过《卫生系统证据》（Health Systems Evidence）可获得这一主题的数百部证据综合。我们之所以关注证据综合（正如我们在第4.2节中指出的一样），是因为它们使用了系统化的过程，对所有针对同一问题的研究结果进行识别、筛选、评价和综合，从而全面了解已有知识，包括这可能因群体和环境而异。

最多只有一份报告将这些证据形式中的任何一种作为其推荐意见的明确重点。正如第7.1节中提到的，许多报告对数据收集和共享提出了一般性建议，但并未就如何使用数据分析支持决策提出具体建议。

	证据形式	委员会报告频次
描述委员会成员专业知识的依据 (不包括他们的个人简历)	技术评估/成本效果分析	1
	所有其他形式的证据	0
	未明确报告	69
所借鉴的证据来源	建模	13
	证据综合	6
	技术评估/成本效果分析	5
	数据分析	3
	评价	2
	指南	2
	行为/实施研究	1
	定性见解	1
	未明确报告	49
建议的重点	建模	1
	评价	1
	定性见解	1
	行为/实施研究	1
	指南	1
	所有其他形式的证据	0
	未明确报告	66



4.16 质量评价工具的示例（第4.5节的附录）

证据形式	质量评价工具的示例
现存质量评价工具的证据类型	
数据分析	<p>非随机干预性研究偏倚评估工具RoB2 (riskofbias.info) 适用于明确特定因素（包括干预措施）和特定结果相关性的观察性研究，存在以下偏倚风险来源：</p> <ul style="list-style-type: none"> 混杂偏倚（由于一个或多个未被考虑的其他因素，观察到的因素和结果之间的关系与真实关系不同） 选择受试者偏倚 干预措施分类 意向干预偏离偏倚 缺失数据偏倚 结果偏倚 选择性报告偏倚
评价	<p>偏倚风险评估工具（RoB）2 (riskofbias.info) 适用于随机对照试验，混杂偏倚较少，但偏倚风险与上述某些（尽管较少）来源相同：</p> <ul style="list-style-type: none"> 随机化过程中的偏倚 偏离既定干预措施的偏倚 结局数据缺失的偏倚 结果测量的偏倚 选择性报告偏倚
行为/实施研究	请参阅其他部分以了解相关类型的研究或综合
定性研究	<p>JBI定性研究的关键评价清单 (bit.ly/31Lsib1)，有许多非常不同的考虑因素，例如：</p> <ul style="list-style-type: none"> 研究方法与研究问题、数据收集方法、数据呈现和分析、结果解释之间的一致性，以及所述哲学观点与研究方法之间的一致性 研究者的反思性，如从文化和理论上定位研究者的表述，并讨论研究者对研究的影响，反之亦然 研究参与者的代表性及其意见 从分析和数据解释中得出结论的流程
证据综合	<p>证据综合中考虑的相关研究类型见上文</p> <p>系统评价/Meta分析方法学质量评价工具（AMSTAR; amstar.ca）用于评价证据综合的质量，偏倚风险可能来自：</p> <ul style="list-style-type: none"> 通过全面检索已发表的文献和灰色文献，确定所有潜在的相关研究，并且不对语言进行限制 使用有关研究设计和受试者、干预措施/因素、对照和结局的明确标准，筛选出所有解决研究问题的研究，并且至少有两名评价人员应用该标准 对所有纳入的研究进行质量评价和数据提取 对所有纳入的研究结果进行综合 <p>请注意，AMSTAR有两个版本：1) 原始版本，适用于所有类型的综合研究，尽管从领域和条目中删除了一些标准；2) AMSTAR的第二个版本，与随机对照试验的综合更为相关</p> <p>推荐分级的评估、制订与评价（GRADE; bit.ly/3C9pMrx），用于评价干预结局的证据质量，其中影响证据质量的因素有：</p> <ul style="list-style-type: none"> 降级因素：偏倚风险（来自随机对照试验的证据起始质量为高，来自观察性研究的证据初始质量为低，随后根据RoB2或ROBINS-I进行升降级）、不精确性（例如，一两个小型研究）、不一致性（例如，两项研究显示结果差距很大）、间接性（例如，使用替代措施或研究的非典型环境）、以及发表偏倚（例如，由于缺乏研究注册平台，此偏移在观察性研究中更常见；或由于发表阳性结果研究的商业动机，此在企业资助的研究中更常见） 升级因素：大效应量、存在剂量-效应关系，以及负偏倚 <p>GRADE CERQual (cerqual.org) 用于对感兴趣的现像做定性描述的证据分级，其中：</p> <ul style="list-style-type: none"> 由于对方法学局限性的担忧（因使用类似上述JBI的关键评价工具发现研究设计或报告方式上的问题）、相关性（因主要研究的环境与综合问题的环境有实质性的不同）、结果一致性（因一些数据与研究结果相矛盾或含糊不清）和数据充分性（因数据不充分或仅来自少数研究或受试者），证据质量则被降级



技术评估/成本效果分析	<p>国际卫生技术评估机构协作网 (International Network of Agencies for Health Technology Assessment, INAHTA) 技术评估质量清单 (bit.ly/2YJVMK)，14个条目中有两个条目涉及证据综合的方法（类似于AMSTAR），另一个条目涉及评估是否通过附带的成本效果分析（与当地，即国家或次国家的成本核算数据）进行适用性分析，以及考虑当地的法律、伦理和社会影响</p> <p>成本效果分析的德拉蒙德 (Drummond) 清单 (bit.ly/3FbnB8R)，以及更普遍的经济评价，包括研究设计、数据收集以及结果分析和解释的问题</p> <p>飞利浦 (Philips) 成本效果分析清单包括决策分析模型部分 (bit.ly/3FcWBGc)，其中包含涉及模型的结构（例如，明确的理由、合理的假设和适当的时间范围）、采用的数据（例如，来自观察研究的基线概率、来自随机对照试验的治疗效果以及对四种不精确性的评估，即模型的结构、遵循的方法步骤、所研究人群的异质性和使用的参数），以及一致性（内部和外部）。TRUST工具同时可以来评价决策分析模型的不精确性 (bit.ly/3quFSKp)</p>
指南	<p>AGREE II工具 (bit.ly/30qyFAb) 用于评价指南的制订、报告和评价（或质量评价），该工具包含6个领域的23个条目，每个领域独立评分：</p> <ul style="list-style-type: none"> 描述范围和目的 利益相关者（公民/患者和专业人士）的参与 制订严谨性（将证据综合作为一种证据，使用严格的推荐意见形成过程，并将推荐意见与支持性证据联系起来） 表达清晰性 适用性 编辑独立性（与资助者和小组成员的利益冲突有关） <p>GRADE (bit.ly/3C9pMrx) 用于评估推荐意见的强度，它有四个关键考虑因素：</p> <ul style="list-style-type: none"> 平衡（权衡）理想和不理想的结果，考虑到对理想和不理想结果效应量的最佳估计，以及对这些结果的重要性（评估偏好和价值观） 干预措施对重要结局效应量的把握度（见前一行） 使用偏好和价值观及可变性资源的把握度

暂无质量评价工具的证据类型



建模	<p>对于大多数类型的模型，目前还没有广泛接受使用的工具，然而，存在一些关于模型的共性问题（就像上述飞利浦清单中部分列出的问题），例如：</p> <ul style="list-style-type: none"> 模型的结构（例如，明确的理由、合理的假设和适当的时间范围） 使用的数据（例如，来自观察性研究的基线概率，来自一系列来源的干预效果*，以及对四种不可靠性的评估，即模型的结构、遵循的方法步骤、研究人群的异质性以及使用的参数） 一致性（内部和外部） 软件或工具的可用性，便于其他人可以对其进行评价 <p>*COVID-19面临的挑战之一是，通常用于获取干预效果的研究设计，如随机对照试验，在伦理上或后勤上存在困难和（或）需要时间来完成，因此需要使用其他研究设计，并需要征求专家意见（并且有一些系统和透明的方式进行，如SHELF-见 bit.ly/30nteC4）</p>
----	--

某些暂无质量评价工具的证据类型所使用的方法



人工智能	暂无广泛认同的工具
------	-----------

4.17 参考文献

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.