THEORETICAL FRAMEWORKS AND TOOLS TO SUPPORT PROCESSES FOR FUTURE HEALTHCARE ECOSYSTEMS

MICHIEL SEBASTIAAN OERBEKKE

Theoretical frameworks and tools to support processes for future healthcare ecosystems

Michiel Sebastiaan Oerbekke

Theorerical frameworks and tools to support processes for future healthcare ecosystems
Michiel Sebastiaan Oerbekke
PhD thesis, Utrecht University, the Netherlands

ISBN/EAN: 978-90-393-7678-2
DOI: https://doi.org/10.33540/2337
Printed by: IPSKAMP Printing, Enschede, the Netherlands

Financial support for printing this thesis was kindly provided by the Knowledge Institute of the Dutch Association of Medical Specialists and Cochrane Netherlands.

Theoretical frameworks and tools to support processes for future healthcare ecosystems

Theoretische raamwerken en hulpmiddelen ter ondersteuning van processen voor toekomstige gezondheidszorgecosystemen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 11 juni 2024 des ochtends te 10.15 uur

door

Michiel Sebastiaan Oerbekke

geboren op 8 september 1988
te Almelo

Promotor:
Prof. dr. L. Hooft

Copromotor:
Dr. M.J. van der Laan

Beoordelingscommissie:
Prof. dr. R. de Bree MD (voorzitter)
Prof. dr. H. Gardarsdottir
Prof. dr. M.L. Bots
Prof. dr. D.M.J. Delnoij
Prof. dr. P.I. Spuls

CONTENTS

CHAPTER 1	General introduction	7
CHAPTER 2	Converging the evidence ecosystem and learning healthcare system to speed up and improve the complex processes from evidence generation to implementation and learning in clinical practice	17
CHAPTER 3	A whole system approach for evaluating best practices in healthcare	153
CHAPTER 4	Designing tailored maintenance strategies for systematic reviews and clinical practice guidelines using the Portfolio Maintenance by Test-Treatment (POMBYTT) framework	191
CHAPTER 5	Priority-setting for resource constraints in systematic review or clinical practice guideline maintenance strategies: Extending the Portfolio Maintenance by Test-Treatment (POMBYTT) framework	259
CHAPTER 6	Introducing re-weighted range voting in clinical practice guideline prioritization: development and testing of the RE-Weighted Priority-Setting (REPS) tool	301
CHAPTER 7	Exploring transparent reporting and data availability in systematic reviews to identify subgroup evidence: imaging for suspected hepatocellular carcinoma in the non-cirrhotic liver	401
CHAPTER 8	Data sources and methods used to determine pretest probabilities in a cohort of Cochrane diagnostic test accuracy reviews	413
CHAPTER 9	General discussion	449
APPENDICES	Summary Samenvatting Dankwoord About the author	473

CHAPTER 1
General Introduction

INTRODUCTION

In 2013 the Committee on the Learning Healthcare System in America of the Institute of Medicine of the National Academies stated:

> *"If home building were like healthcare, carpenters, electricians, and plumbers each would work with different blueprints with very little coordination."* [1]

This quote suggests a lack of coordination within the entire healthcare system as a whole. Recognizing this, the Institute of Medicine therefore organized several round table meetings to describe and explore a Learning Healthcare System and its associated challenges.[1-7] With describing the Learning Healthcare System, the Institute of Medicine aimed to establish a system where 90% of clinical decision-making was supported with accurate and up-to-date clinical information based on the best-available evidence.[2] Consequently, such a system would be designed to generate and apply the best available evidence to provide the most appropriate care to the patient.[2] Care always starts with providing proven best-practices,[3, 4] as they form the foundation of the delivery of care. The interaction between patients and clinicians in the delivery of care generates valuable data, often captured in electronic health records. These data contribute to the ongoing improvement of care and innovation within the Learning Healthcare System. This process is facilitated by lining up research, informatics, incentives, and culture.[3] Within the Learning Healthcare System-framework, three types of learning can be identified: 1) exchanging information between the clinical and research domain, 2) iteratively converting data to knowledge, knowledge to actions, and actions to data, and 3) recurring interactions with stakeholders.[8]

While the Learning Healthcare System is primarily focused on providing care and learning from these interactions, the process for developing accurate and up-to-date products (e.g., systematic reviews and guidelines) based on the best-available primary evidence do not appear to be fully integrated in this system. Despite acknowledging the connection between practice and primary research,[8] the evidence synthesizers and translators seem to be less explicitly involved (besides from describing an entry for learning from external evidence[9]). Instead, the primary data generation, synthesis, and translation process can be described as an Evidence Ecosystem,[10-12] where evidence is developed and contextualized rather than being centered around care provision. Within this Evidence Ecosystem, each participating organization probably has its own role, for example, research organizations contribute to primary data generation, evidence synthesis organizations undertake data syntheses, and guideline developing organizations develop clinical practice guidelines. As primary research is included in systematic reviews, which in turn are included in guidelines, and guidelines are disseminated to clinical practice, the Evidence Ecosystems is a complex connected system where data and products interlace. Therefore, the coordination and collaboration required for evidence generation and contextualization extend beyond Learning Healthcare Systems to encompass Evidence Ecosystems as well. Despite the importance of infrastructure and collaboration for the flow of evidence to clinical care,[12] efforts within evidence production, synthesis, and translation appear to be siloed.[11] Connecting Learning Healthcare Systems

and Evidence Ecosystems, as part of a future healthcare ecosystem, could involve creating a larger network of organizations evaluating, learning, collaborating, and co-creating products in their own context and with their own needs.

As products circulate within the future healthcare ecosystem, it is likely that the original developer remains responsible for these products. For example, the evidence synthesizer remains responsible for their systematic reviews as they disseminate throughout the ecosystem. This accountability extends to monitoring the status of these systematic reviews over time. As new evidence is published in an increasing rate,[13] systematic review might become outdated. In fact, it is suggested that 15% of systematic reviews may be outdated within a year after publication.[14] Similar responsibilities apply to clinical practice guideline developers, where approximately 8% of the guideline recommendations may become outdated in the first year after development.[15] Changes in what are considered to be important outcomes or changes in the available interventions could require updates.[16] This potentially leaves conclusions in systematic reviews and recommendations in guidelines sub-optimal or even harmful, causing patients to potentially receive outdated or unnecessary care. It underscores the importance of continuously updating and refreshing products within the future healthcare system as needed. Consequently, both evidence synthesizers and translators are responsible to continuously update their circulating products. This becomes vital when these products interlace, as systematic reviews are often substantiate clinical practice guidelines, informing health-care decisions at the point of care. Furthermore, evidence translators may have additional responsibilities in presenting synthesized and translated knowledge into easily understandable formats so that end-users at the point of care can readily understand and apply the information provided.

Looking forward, such roles and responsibilities within a future healthcare ecosystem also may require participating organizations to adjust their current processes to better align with these evolving demands. These revised processes should aim to ensure timely, accurate, and up-to-date support for clinicians and patients at the point of care. For example, such adaptations may involve implementing processes to support continuous learning at the point of care, as well as supporting the ongoing developing and updating of syntheses and clinical practice guidelines. Additionally, there is a need to translate evidence into interpretable formats for broader utilization. Herein we must acknowledge that resources for such processes can be limited. Therefore, the development of new methodological guidance for processes and tools may be required to support organizations moving towards a future healthcare ecosystem while dealing with these resource limitations. Failing to contemplate, develop, and pilot these processes and tools for the future healthcare ecosystem, we remain to be carpenters, electricians, and plumbers building a home with different blueprints. A further complicating factor is the diverse contexts that organizations operate in, each potentially requiring specific adaptations to meet their needs. Ideally, comprehensive frameworks and tools for such processes are developed to enable organizations tailoring processes to their own unique needs and contexts.

The work in this thesis was established from an interdisciplinary perspective at the cross-section of methodology, organizational practices, evidence synthesis, and evidence translation. It represents a collaborative endeavor of the Knowledge Institute of the Dutch

Association of Medical Specialists and Cochrane Netherlands.

The Knowledge Institute of the Dutch Association of Medical Specialists consults and supports developing, applying, and training in quality products and policies in the medical specialty domain, for example, with clinical practice guideline development, visitation training, and evidence syntheses.[17] Through the translation of synthesized evidence to the Dutch context, clinical practice guidelines serve as the cornerstone of healthcare provision for medical specialists. However, the development of clinical practice guidelines may face challenges due to limited evidence availability, preventing making strong clinical care policies in terms of certainty and/or direction. This means that current guideline recommendations may be substantiated on evidence in which there is a (very) low certainty that their results reflect the truth for the average patient. It may even occur that there is no scientific evidence at all to base a recommendation on. The Knowledge Institute of the Dutch Association of Medical Specialists believes in evaluating current practice to substantiate the current clinical practice and refining clinical practice guidelines by prioritizing knowledge gaps for primary research communities.[18] This, however, requires robust processes and tools for the clinical practice guideline portfolio to accommodate for an evolving evidence base and using priority-setting to address resource constraints effectively.

The Cochrane Collaboration is an independent, global organization that produces trusted and timely evidence syntheses to address and support the most important health decisions.[19] Over 9000 of their Cochrane reviews are available in the Cochrane Library,[20] including reviews of diagnostic tests, prognosis studies, and interventions. Besides producing evidence syntheses, the Cochrane Collaboration develops new methods for such evidence syntheses.[19] Representing the Cochrane Collaboration's mission and vision, Cochrane Netherlands operates from the Netherlands as one of the global geographical groups. Such groups serve various functions, such as supporting Cochrane review authors, performing commissioned evidence syntheses, providing methodological consultancy, offering educational courses, and engaging in methodological research in the areas of evidence-based practice and systematic reviews.[21] As an integral part of the global evidence synthesis community, Cochrane Netherlands could play an important role in a future healthcare ecosystem. The rigorous and trustworthy systematic reviews and commissioned evidence synthesis reports produced by Cochrane (Netherlands) are used in Dutch medical specialists' clinical practice guidelines for a wide variety of topics. The development of these Dutch guidelines was supported by the Knowledge Institute of the Dutch Association of Medical Specialists. For example, Cochrane reviews have substantiated guidelines concerning antenatal corticosteroid use before an elective c-section,[22, 23] clinical tests for sleep apnea in adults,[24, 25] and deep brain stimulation for epilepsy.[26, 27] This shows the pivotal role of systematic reviews substantiating clinical practice guidelines to inform important healthcare decisions by clinicians and patients at the point of care. Consequently, it underscores the need to keep both systematic reviews and clinical practice guidelines maintained over time to ensure their ongoing relevance and utility for clinicians and patients at the point of care.

By introducing the Knowledge Institute of the Dutch Association of Medical Specialists and the Cochrane Collaboration together with their respective products, a critical connection between the evidence translation and evidence synthetization is established.

This linkage plays a pivotal role in supporting important health decisions at the point of care. Looking ahead, the potential for these connections and collaborations to deepen is evident as networks formalize and develop into a comprehensive healthcare ecosystem. Even within this ecosystem, organizations are likely to have different roles and carry unique responsibilities. Adaptation of their processes is probably essential to serve these roles and responsibilities effectively in a future ecosystem, encompassing tasks such as continuous product maintenance and translating evidence in understandable formats. Consequently, organizations forming links and networks may find it beneficial to implement frameworks and tools fostering the transition towards a future healthcare ecosystem.

As we envision a future healthcare ecosystem, it becomes evident that organizations face the imperative of adapting their processes and responsibilities to align with the evolving healthcare landscape. This adaptation is crucial for ensuring that clinicians and patients receive timely, accurate, and up-to-date support at the point of care. The forthcoming chapters of this thesis provide theoretical frameworks and tools for processes regarding specific responsibilities required for organizations to thrive within this future healthcare ecosystem.

THESIS OUTLINE

This thesis aims to provide both theoretical frameworks and tools to support real-world processes in a healthcare ecosystem. Figure 1 projects the thesis' chapters on entities in a potential future healthcare ecosystem.

The scientific literature outlines two distinct systems: the learning healthcare system[2, 3, 8, 12, 28, 29] and the evidence ecosystem.[10-12] The learning healthcare system focuses on internal learning and improvement within the clinical field through activities such as analyzing clinical data.[29] On the other hand, the evidence ecosystem concept involves the interactions between the primary research, the evidence synthesis, and evidence translation.[10, 11] In real-world practice, processes in both systems may be interconnected and mutually beneficial, with the success of each system potentially dependent on the interactions between both systems. **Chapter two** presents a synthesized model that visualizes and explores how the learning healthcare system and evidence ecosystem could operate as one symbiotic healthcare ecosystem as a framework to speed up and improve complex processes from evidence generation to implementation and learning.

An important part of the symbiotic healthcare ecosystem is the point of care, where the essential patient-clinician interaction takes place. As earlier stated, the foundation of care always lies in providing best-practices.[3, 4] Doctors, nurses, healthcare managers and supporting staff consistently implement and refine their practices, maintaining a commitment to excellence. Importantly, these best practices continuously learn, both from their internal processes and external information.[9] **Chapter three** introduces a comprehensive framework featuring multiple indicators and a tool to evaluate a best practice. This may serve as valuable mirroring information, allowing practices within the healthcare ecosystem to reflect upon and learn from one another.

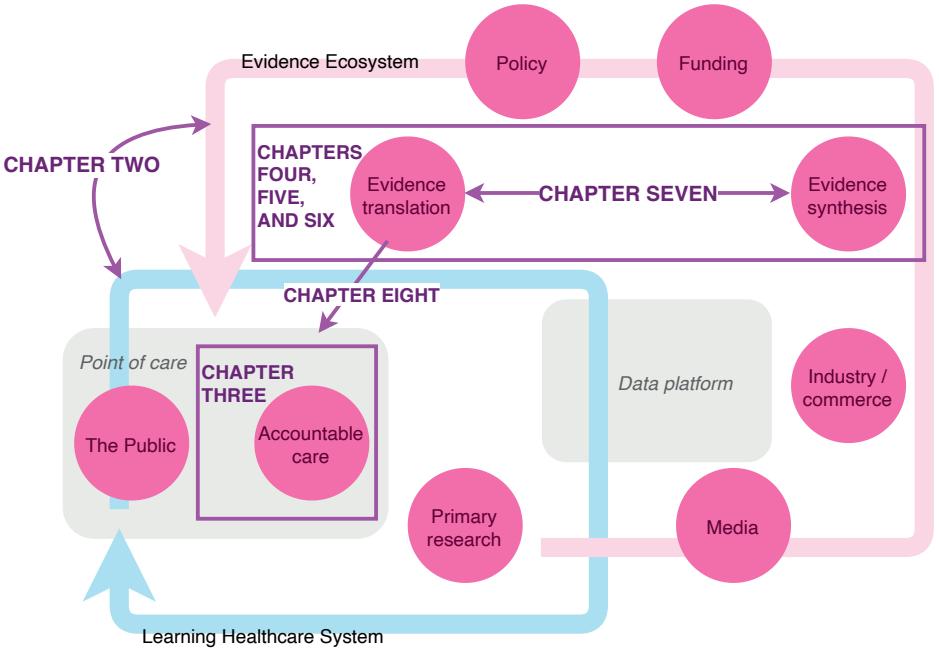


Figure 1. The thesis' chapters projected on a schematic future healthcare ecosystem. *The pink circles represent entities in a future healthcare ecosystem. The Learning Healthcare System (soft blue arrow) and the Evidence Ecosystem (soft pink arrow) are schematically depicted. The bright pink arrows indicate chapters focused on processes or tool to establish connections in the system (Chapters two, seven, and eight). Chapters related to frameworks and tool for processes for one or multiple entities in a future healthcare system are denoted by bright pink boxes (Chapters three, four, five, and six).*

As a learning healthcare system is designed to apply best evidence,[2] the roles of evidence synthesizers and evidence translators become important in the symbiotic healthcare ecosystem. They are responsible to develop crucial products, such as comparative effectiveness systematic reviews and clinical practice guidelines. As these products circulate in the symbiotic healthcare system and serve as best evidence, a continuous effort to refresh and update these products to reflect the latest state of science and practice is imperative. The work described in **Chapter four** introduces a framework in analogy to a test-treatment strategy to design or tailor a process to maintain portfolios of systematic reviews or clinical practice guidelines. This framework aids in designing a maintenance strategy, providing organizations with multiple options to manage systematic reviews and clinical practice guidelines in their portfolio, irrespective of whether these products are outdated or not.

Unfortunately, resources such as time, budget, and workforce are constantly limited. This also applies to the available resources for the continuous maintenance of systematic reviews and clinical practice guidelines. Therefore, organizations must discern which systematic reviews and clinical practice guidelines have the highest priority for resource allocation. **Chapter five** builds upon the framework described in **Chapter four** by incorporating theoretical concepts of priority-setting assessments into maintenance strategies. This extension distinguishes the constructs of 'assessing the need for updating' from 'priority-setting', utilizing concepts from diagnostic test accuracy methodology in analogy to describe the roles of priority-setting assessments.

By recognizing the importance of priority-setting in maintenance strategies within the constraints of limited resources, it becomes important to understand the composition of these priority-setting assessments and how to assess priority itself. **Chapter six** first provides a theoretical frame of reference to help understand the components of a priority-setting assessment. Subsequently, the chapter introduces a tool integral to priority-setting assessments, aimed at evaluating the priority of items such as key questions or guideline sections. The tool is designed to boost less represented perspectives among participants in a priority-setting assessment if deemed necessary and to seamlessly integrate into maintenance strategies described in **Chapter four** and **five**. Importantly, the tool is flexible, allowing organizations the autonomy to self-tailor their procedures, as long as the input format aligns with the tool's requirements (i.e., a matrix of one score per participant per item).

When developing new products or updating existing ones within the symbiotic healthcare ecosystem post priority-setting, there's a growing need for efficiency by reusing efforts already invested by other entities in the system. However, this heightened efficiency underscores the need for transparency and completeness of products, and increased availability and accessibility of data. **Chapter seven** bridges the evidence synthesis and translation entities by illustrating how information and data from systematic reviews can be effectively reused by guideline developers. Using a real-world example in the field of hepatocellular oncology, this chapter demonstrates benefits of efficiently reusing information and data from one entity by another entity. This approach seems particularly helpful during resource-constrained scenarios in the development of a clinical practice guideline segment.

Another challenge may arise when there are products circulating within the symbiotic healthcare ecosystem reflecting the latest state of practice and science but lack interpretability for their end-users. For example, understanding the practical implications of a diagnostic test's summary sensitivity of 90% and a specificity of 80% for clinicians and patients at the point of care can be daunting. **Chapter eight** focuses on establishing a connection from evidence translators to end-users at the point of care, ensuring that diagnostic evidence is adequately translated into an understandable and interpretable format. This chapter provides considerations and guidance for communicating diagnostic test accuracy parameters in an interpretable manner using a hypothetical cohort, specifically aiding the decision-making process for selecting a representative pre-test probability.

Chapters two to **eight** collectively describe theoretical frameworks and tools for supporting processes throughout a future healthcare ecosystem. This thesis concludes with a general discussion in **Chapter nine,** providing a perspective on **Chapters three** to **eight** in light of the insights presented in **Chapter two**.

LITERATURE LIST

1. Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
2. Institute of Medicine. The Learning Healthcare System: Workshop Summary. In: Olsen L, Aisner D, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright © 2007, National Academy of Sciences.; 2007.
3. Institute of Medicine. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. In: Grossmann C, Powers B, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
4. Institute of Medicine. Patients Charting the Course: Citizen Engagement and the Learning Health System: Workshop Summary. In: Olsen L, Saunders RS, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
5. Institute of Medicine. Large Simple Trials and Knowledge Generation in a Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
6. Institute of Medicine. Genomics-Enabled Learning Health Care Systems: Gathering and Using Genomic Information to Improve Patient Care and Research: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2015 by the National Academy of Sciences. All rights reserved.; 2015.
7. Medicine Io. Observational Studies in a Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
8. de Bruin J, Bos C, Struijs JN, Drewes HW, Baan CA. Conceptualizing learning health systems: A mapping review. Learn Health Syst. 2023;7(1):e10311.
9. Guise JM, Savitz LA, Friedman CP. Mind the Gap: Putting Evidence into Practice in the Era of Learning Health Systems. J Gen Intern Med. 2018;33(12):2237-9.
10. Cartabellotta A, Tilson JK. The ecosystem of evidence cannot thrive without efficiency of knowledge generation, synthesis, and translation. J Clin Epidemiol. 2019;110:90-5.
11. 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. J Clin Epidemiol. 2020;123:153-61.
12. Vandvik PO, Brandt L. Future of Evidence Ecosystem Series: Evidence ecosystems and learning health systems: why bother? J Clin Epidemiol. 2020;123:166-70.
13. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326.
14. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147(4):224-33.

15. Martínez García L, Sanabria AJ, García Alvarez E, Trujillo-Martín MM, Etxeandia-Ikobaltzeta I, Kotzeva A, et al. The validity of recommendations from clinical guidelines: a survival analysis. Cmaj. 2014;186(16):1211-9.
16. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? Bmj. 2001;323(7305):155-7.
17. Kennisinstituut van de Federatie Medisch Specialisten. Expertise. 2023. Available from: https://demedischspecialist.nl/kennisinstituut/expertise [Accessed on: 3 August 2023]
18. Kennisinstituut van de Federatie Medisch Specialisten. Zorgevaluatie. 2023. Available from: https://demedischspecialist.nl/kennisinstituut/expertise/zorgevaluatie [Accessed on: 3 August 2023]
19. Cochrane. About us. 2023. Available from: https://www.cochrane.org/about-us [Accessed on: 3 August 2023]
20. Cochrane. Search reviews (CDSR). 2023. Available from: https://www.cochranelibrary.com/cdsr/reviews [Accessed on: 3 August 2023]
21. Cochrane Netherlands. Over ons. 2023. Available from: https://netherlands.cochrane.org/over-ons [Accessed on: 3 August 2023]
22. Nederlandse Vereniging voor Obstetrie en Gyneacologie. Dreigende vroeggeboorte: Corticosteroiden bij geplande sectio tussen 37+0 en 39+0 weken zwangerschapsduur. 2022. Available from: https://richtlijnendatabase.nl/richtlijn/dreigende_vroeggeboorte/corticosteroiden_bij_geplande_sectio_tussen_37_0_en_39_0_weken_zwangerschapsduur.html [Accessed on: 3 August 2023]
23. Sotiriadis A, McGoldrick E, Makrydimas G, Papatheodorou S, Ioannidis JP, Stewart F, Parker R. Antenatal corticosteroids prior to planned caesarean at term for improving neonatal outcomes. Cochrane Database Syst Rev. 2021;12(12):Cd006614.
24. Nederlandse Vereniging van Artsen voor Longziekten en Tuberculose, Nederlandse Vereniging voor Keel-Neus-Oorheelkunde en Heelkunde van het Hoofd-Halsgebied. Obstructief slaapapneu (OSA) bij volwassenen: Klinische tests bij diagnose OSA. 2022. Available from: https://richtlijnendatabase.nl/richtlijn/osa_bij_volwassenen/het_concept_osa_actuele_inzichten/klinische_testen_voor_osa.html [Accessed on: 3 August 2023]
25. Heus P, Damen JAA, Japenga EJ, Spijker R, Scholten RJPM, Hooft L. Diagnostiek bij patiënten met slaperigheidsklachten en verdenking op slaapapneu (OSA). Cochrane Netherlands; 2018.
26. Nederlandse Vereniging voor Neurologie. Epilepsie: Diepe hersen stimulatie bij epilapsie. 2020. Available from: https://richtlijnendatabase.nl/richtlijn/epilepsie/neurostimulatie_bij_epilepsie/diepe_hersen_stimulatie_bij_epilepsie.html [Accessed on: 3 August 2023]
27. Sprengers M, Vonck K, Carrette E, Marson AG, Boon P. Deep brain and cortical stimulation for epilepsy. Cochrane Database Syst Rev. 2017;7(7):Cd008497.
28. Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26(2):w107-18.
29. Easterling D, Perry AC, Woodside R, Patel T, Gesell SB. Clarifying the concept of a learning health system for healthcare delivery organizations: Implications from a qualitative analysis of the scientific literature. Learn Health Syst. 2022;6(2):e10287.

CHAPTER 2

Converging the evidence ecosystem and learning healthcare system to speed up and improve the complex processes from evidence generation to implementation and learning in clinical practice

M.S. Oerbekke
F.M. Janssen
M.J. van der Laan
L. Hooft

In preparation

ABSTRACT

BACKGROUND

Literature has traditionally discussed the concepts of evidence ecosystems and learning healthcare systems independently. Recent developments have highlighted potential benefits of integrating these systems. Converging both systems in as a single symbiotic framework can help address and overcome challenges within both individual systems. By combining these two systems, organizations may understand and support efficient generation, implementation, and application of relevant evidence in a timely manner. Therefore, the aim of this study was to integrate the EES and LHS into a single symbiotic model, serving as a framework to speed up and improve the complex processes from evidence generation to implementation and learning in clinical practice.

METHODS

MEDLINE was searched to identify studies, reports, and reviews describing characteristics of an evidence ecosystem, learning healthcare system or both. Extracted characteristics from both systems were merged, qualitatively analyzed for emerging themes, and then visualized. These analyses and visualizations were used to construct the symbiotic healthcare ecosystem model as a framework converging the EES and LHS.

RESULTS

The symbiotic healthcare ecosystem model comprises of various entities. Co-creation with stakeholders ensures the relevance of data and products, while quality actions are essential for fostering trust in the data and products circulating the system. This framework has two key locations: 1) the point of care, where real-world data is generated during the patient-clinician interaction, and 2) the data platform, a digital location facilitating interoperable data exchange, products, and knowledge throughout the system. An example is provided to illustrate how the involved entities collaborate within this model to generate, develop, utilize, and evaluate products across the system.

CONCLUSION

To capitalize on a symbiotic healthcare ecosystem, it seems important to understand vital aspects for connecting the sub-systems. The presented framework serves as a blueprint for understanding complex processes across both sub-systems. Examining these complex processes through the framework's lens can assist in identifying both necessary and unnecessary delays, including inefficient organizational processes, along the pathway from evidence generation to implementation and learning.

Keywords: Learning healthcare system, evidence ecosystem, symbiotic healthcare ecosystem, concept formation, theoretical model

1. INTRODUCTION

The Institute of Medicine stated in 2007 that a limited capacity for the timely generation of evidence was progressively responsible for delivered unimportant care and important undelivered care, besides failure to apply best evidence.[1] Furthermore, there seems to be a delay of various durations in the health research translation process to benefit patients.[2, 3] One of the problems is that research, evidence syntheses, clinical practice guideline (CPG) production, and evidence adoption are seemingly performed in siloes.[4] An evidence ecosystem (EES) should bridge the gap between the siloes of primary research, the evidence syntheses, and CPG development.[4] Here, the Digital and Trustworthy EES framework proposes a closed loop of evidence production, evidence synthetization, CPG development, dissemination, implementation, and evaluation and improvement.[5]

At the same time, a learning healthcare system (LHS) is a complementary framework to an EES.[5, 6] An LHS is aimed at analyzing clinical data, generating evidence, and translating evidence into practices and enabling continuous quality improvements, innovations, and organizational learning.[7] Interestingly, one of the conceptual EES frameworks describes evaluation and improving practices,[5] while the LHS is described generating evidence[1, 7] and learning from critically appraised external evidence.[8] Given that both concepts seem to create an entry for each other's processes and products, they can potentially function as one system together. The EES and LHS likely engage in mutually beneficial interactions, forming what could be perceived as a symbiotic healthcare ecosystem (SHE). Certain problems and challenges need both the EES and LHS as sub-systems of a SHE to be effectively addressed. As research, synthesis, translation and adoption initiatives are largely siloed,[5] the problem of timely implementation of new evidence in an LHS might (partially) be attributed to these inefficiencies and a lack of clinically relevant products in the EES. For example, barriers for implementing CPGs, such as their clarity, doubts about the efficacy of interventions and clinical outcomes, the sentiment that their evidence is incorrect or insufficient, or disagreement with the recommendations,[9] might cohere with a lack of such relevance.

To understand and support efficient generation, implementation, and application of relevant evidence in a timely manner, it is important to explore the interplay between the EES and LHS as interdependent sub-systems in a symbiotic model. This entails a focused examination on data generation, evidence product development, and the seamless exchange of these data and products within and between the EES and LHS. While the EES and LHS were previously described separately, it was recently suggested to integrate EES advancements into the LHS.[6] However, understanding and setting-up efficient complex real-world processes in the pathway from evidence generation to implementation and learning, spanning both the EES and LHS, may require more direction for guidance than solely integrating new developments from complementary systems. Therefore, an important next step is to integrate the EES and LHS into one SHE framework.

The envisioned SHE should establish a framework wherein entities seamlessly communicate, effortlessly exchange, and continuously learn. Potentially relevant stakeholders, pathways,

and infrastructure involved in new or existing complex real-world processes across sub-systems might be identified through the SHE-framework, together with the required data and products. A symbiotic framework could thereby aid in understanding inefficiencies along such complex processes. Therefore, the aim of this study was to integrate the EES and LHS into a single symbiotic model, serving as a framework to speed up and improve the complex processes from evidence generation to implementation and learning in clinical practice.

2. METHODS

2.1 SEARCH STRATEGY AND STUDY SELECTION

On the 2nd and 12th of April 2022 searches were performed in MEDLINE via PubMed to identify reports describing an EES and LHS, respectively. The search string and the selection criteria can be found in Additional File 1. Given the extensive yield of studies in the initial search strategy for LHS, the search results were confined to (systematic) reviews and reports to avoid duplicative efforts with other published reviews on LHS characteristics. Two authors (MSO, FMJ) independently screened the titles and abstracts of the search retrieval for potentially relevant reports and articles, with conflicts resolved in a meeting. Thereafter, all potentially relevant articles and reports were read full text for their eligibility by two authors (MSO, FMJ) blinded from each other's selection decisions, resolving any conflicts in a subsequent meeting. Additionally, one author (MSO) hand-searched the reference lists of the included studies for potentially relevant articles missed by the search strategy.

2.2 DATA EXTRACTION

Data was extracted by one author (MSO) and checked by a second author (FMJ). Data regarding the following EES or LHS characteristics were extracted: the areas or components in the system, the actors, links or cooperation in the system, products in the system, data handling, data quality and values or procedures, governance, and use of (digital) platforms. For EESs, methods for product maintenances were additionally extracted. For LHSs, data sources and methods or mechanisms for learning or evaluation were additionally extracted.

2.3 SINGLE SYSTEMS, INTERMEDIARY ANALYSES, AND MODEL DEVELOPMENT

Distinctive characteristics extracted from the EES and LHS datasets were summarized to delineate each system individually. Subsequently, the extracted items from both systems were combined per characteristic of interest where possible. The items were qualitatively clustered by one author (MSO) to identify emerging themes within each characteristic, serving as a first intermediary qualitative analysis. Extracted items within each characteristic were immediately treated as code labels, which were grouped together to directly derive themes from the data within each characteristic by hand in text-processing software. Identical or similar code labels (e.g. electronic health records or similar terms) from different references

were merged into a single item for data reduction. This intermediate analysis was checked by and discussed with a second author (FMJ) to ensure alignment and consensus, whereafter MSO adjusted the analysis accordingly. Following the thematic grouping of data, one author (MSO) visualized potential links, mechanisms, relations, and hierarchies per characteristic as thematic visualizations from the derived themes and extracted items as subsequent intermediary analyses in flowchart design software. These visualizations were iteratively discussed with a second author (FMJ) for alignment and consensus in multiple meetings to further refine the understanding. All intermediary analyses were used to construct the SHE-model layers that encapsulated different aspects of a SHE, resulting in an aggregated model. A complex process spanning both the EES and LHS was mapped to the aggregate model providing an example to illustrate the application of the SHE framework as a proof of concept.

3. RESULTS

3.1 SEARCH AND DATA EXTRACTION

Thirty-six unique articles, reviews and reports were selected for EES[4, 5, 10-24] and LHS[5, 7, 8, 25-41] characteristics, respectively (Figure S1 in Additional File 2). One article [5] described characteristics of both systems. Excluded studies in the full-text selection phase with reasons for exclusion can be found in Table S1 in Additional File 2. The extracted data for both the EES and LHS can be found in Tables S1 and S2 in Additional File 3, respectively.

3.2 SUMMARIZING INDIVIDUAL SYSTEMS

Table 1 highlights key characteristics of EESs and LHSs based on the extracted data. Both systems seem to include patients, clinicians, researchers, and policymakers, although EESs appear to prioritize roles related to generating, synthesizing, and translating within their framework. These EESs seem to focus on reducing research waste and generating trustworthy products. Furthermore, LHSs seems to prioritize roles regarding evidence utilization, evaluation, and learning while seemingly placing greater emphasis on including patients and their family members. Researchers in LHSs seem to produce primary evidence (both observational and trials) based on electronic health records, patient reported outcome measures, and experiences of care during routine care. Such LHSs appear to emphasize using these products for continuous evaluation and learning within their framework.

3.3 INTERMEDIARY ANALYSES

Emerging themes from the qualitative clustering of EES and LHS data per characteristic are described in Table S3 in Additional File 3. The visualizations based on the emerging themes and clustered data per characteristic (except for '*links and cooperation*', as this is visualized in the model's foundation layer) can be found in Figures S1-S10 in Additional File 3.

Table 1 – Highlights of key system characteristics

			Within the Evidence Ecosystem Framework	Within the Learning Healthcare System Framework
Public and patients			The public[14, 21] Patients[4, 5, 14, 19]	Communities[7, 30, 35, 39] Patients[5, 7, 25-33, 35, 39, 41] Family-members[7, 25, 27, 30-32, 35]
Care		*Actors*	Clinicians, health care professionals or providers, physicians, (general) practitioners, or practitioner-academics[4, 5, 11-14, 19, 21, 23]	Clinician-scientists[25] Clinicians, physicians, or health care providers[5, 7, 26-36, 39, 41]
		Products / data in the system	Electronic health record data[5, 20] Real-world data[4, 5]	Best-practices[35] Medication alerts[35] Data-dashboards[31] Registries[34] Nationwide health information network / exchange[34] Patient data and electronic health records[5]
Evidence	*Primary research*	*Actors*	Producers of information/ knowledge[15, 17-19, 21, 24] Researchers, trialists, or academics[4, 11-14, 17, 20, 24]	Scientific community, research networks, or quality and research agencies[5, 33, 41] Researchers[7, 26, 27, 29, 31-33, 35, 39, 41]
		Products / data in the system	Primary research, reports, data (e.g. qualitative, quantitative, observational, quasi experimental, longitudinal, trial)[4, 5, 17, 20, 21, 23, 24]	Observational studies (prognostic models,[28, 29, 31, 41] real-world effectiveness[41]) Trials (randomized controlled,[41] pragmatic/simple trials,[28, 29, 40] evaluating models of healthcare delivery,[29] and on patient, unit, or hospital level[29])
	Evidence synthesis	*Actors*	Evidence synthesizers[5, 10, 18, 21]	–
		Products / data in the system	(Living) systematic reviews and evidence syntheses[4, 5, 12-16, 18, 19, 21, 23, 24] (Living) evidence (gap) maps[4, 10, 12, 18, 24]	–
	Evidence translation	*Actors*	Knowledge curators or translators, evidence processors, evidence brokers[17-19, 21, 23, 24] Guideline developers[4, 23]	–
		Products / data in the system	(Living) clinical practice guidelines[5, 13, 16, 19] Policies or evidence briefs[16, 17] Decision-support systems[5, 16].	Clinical practice guidelines[36] Clinical decision support tools/ system[5, 31, 35, 39]
Policy		*Actors*	Policymakers, politicians, or decision-makers[11, 13-15, 17, 18, 21, 23, 24]	Policy makers or Decision makers [31, 33, 41]
		Organizations / departments	Government (departments)[4, 23, 24] Regulatory agencies/authorities[4, 13, 14] Local health authorities[13] Health policy authorities[4]	Federal government[33] Governmental departments[34] Regulatory agencies[34] Federal Health and human services[34]

Support		Methodologists[4, 5, 12] Statisticians[4, 12, 20] (Topic) experts[15, 18, 20] Healthcare system or program managers[4, 13, 21].	Health system leaders[25, 28, 29, 31, 32] Clinical program leaders[26] Network leaders[27] Thought leaders on continuous improvement[31]
Learning	*Evaluation*	–	Networks evaluate their policies, functioning, management structure[27] Evaluate the impact of research resources[27] Evaluate the effects of implementation to redefine or de-implement ineffective or harmful interventions[28] Generate and test hypotheses[33]
	Learning	–	Learning as (intermittent) information exchange between the clinical domain and the research domain[31] Learning as a continuous circular process of converting (routine care) data to knowledge; knowledge to performance; performance to data[31] Learning as a recurrent interaction between stakeholders[31]
Platforms and data sources		Virtual and central or interoperable repository platforms/ repositories[5, 11, 20, 24] Platforms to process or map evidence[5, 21, 24]	Electronic health records[25, 27, 28, 31, 33, 34, 39-41] Patient reported outcome measures[28, 31, 40] Experiences of care (patient or professional)[26, 31] Evidence from trials, studies, and reviews[8]
Values		Reducing research waste[13, 23] Using reporting guidelines[12, 13, 18] Using risk of bias assessments[4, 12] Conflict of interest disclosure and management[13] Promoting best practices in the production, updating, and dissemination of systematic reviews and guidelines[13]	Access and use of data in manners that are socially beneficial[35] Continuously improving quality through knowledge from every care delivery experience[30] Independent researchers publish on the quality of care[26] Standardized processes, protocols and policies (including: communication policies, data-sharing, privacy protection, regulatory compliance)[27] Technology reducing data-entry burden and facilitating care to results in data useful for research and learning[27]

Synthesized from the data-extraction tables. Some items are merged to prevent heterogeneous terms and a wide array of similar items within this table. See Tables S1 and S2 in Additional File 3 for the complete data extraction for both systems.

3.4 CONNECTING BOTH SYSTEMS

Four separate SHE-model layers encapsulate different aspects of a SHE, covering its foundation, its infrastructure, its quality actions, evaluation, and learning aspects, and its governance structure. These four layers are also presented as an aggregated model.

3.4.1 FOUNDATION LAYER

Figure 1 illustrates the foundation layer of the model (using the intermediary analyses of '*links and cooperation*', '*actors*', and '*components, traits and properties*'), serving as the basis for subsequent layers. Within this layer, nine key entities within the SHE are identified: the public, accountable care, primary research, media, industry/commerce, evidence synthesis, policy, evidence translation, and funding. The public and accountable care are involved in co-creation activities throughout the SHE, collaborating in tasks such as co-generating research questions and co-developing products to ensure clinical relevance. Meanwhile, the funding entity plays a crucial role in mobilizing resources for the SHE. The flow of evidence within the system was mapped using push and pull forces of 'evidence pumps'. Evidence users and commissioners create a pull force by requesting evidence, while evidence generators, synthesizers, and translators create a push force by providing evidence and products. Furthermore, two distinct locations within the system are identified in this layer: 1) a physical location, known as the point of care, where interactions occur between the public, accountable care, primary research, and evidence translation; and 2) a digital location represented by the data platform, which facilitates harmonization and exchange of data across the system. For an extended visualization of the model's foundation layer, see Figure S1 in Additional File 4.

3.4.2 INFRASTRUCTURE LAYER

See Figures S2 and S3 in Additional File 4 for extended and simplified visualizations of the infrastructure layer (using the intermediary analyses of '*data sources [LHS only]*', '*platforms*', '*products*', and '*data presentation/packaging*'). Within this layer, it seems that each entity within the SHE contributes to data generation and/or product development, which can be stored in databases accessible through the data platform at the heart of the system for exchange. Various data types were identified throughout the system, including observational data (e.g. citizen monitoring data, real-world data from the point of care, longitudinal data from research, economic evidence), experiences (e.g. patient experiences, qualitative data) and experimental data (e.g. quantitative data an individual patient data from randomized controlled trials). Structured data can be stored in diverse databases, such as biobanks, (trial) registries, clearinghouses, claims databases, health information databases, and marketplace-like databases. These databases are linked and interoperable while accessible through an end-user interface. This interface may offer various tools like catalogues, private research environments, code mappers, interactive data visualizers, and information packaging and synthesis tools. Furthermore, data may be generated to develop new products or be shared with other entities for product development. For example, data from randomized trials are used to develop systematic reviews with meta-analyses, which in turn are used to develop CPGs. The primary research entity, evidence synthesis entity, and evidence translation entity seem to have multiple types of products within the SHE,

including various research types (e.g. prognostic, diagnostic, randomized trials, impact analyses, quasi-experimental), diverse syntheses (e.g. scoping reviews, systematic reviews, summaries, measurement properties reviews, rapid reviews), and contextualized products (e.g. CPGs and decision aids).

3.4.3 QUALITY, EVALUATION, AND LEARNING LAYER

Extended and simplified visualizations of the quality, evaluation, and learning layer (using the intermediary analyses of '*values and procedures for quality*', '*learning and evaluation [LHS only]*', and '*updating/maintenance [EES only]')* in a SHE-model can be seen in Figures S4 and S5 in Additional File 4, respectively. Within this layer it was identified that there is attention for standardization, privacy, security, and transparency for sharing data and products. To maintain ongoing trustworthiness, evidence synthesis and translation entities were identified developing living and rapid products. Various quality actions are employed throughout the symbiotic system to create trust in these processes and products. For example, adopting the GRADE approach (both the GRADE system and GRADE evidence-to-decision), conducting risk of bias assessments, leveraging automated technologies, and adhering to reporting guidelines. These actions seem to aim at generating clear, usable, and relevant evidence that can be trusted by other entities within the SHE. Furthermore, within this layer, learning in the SHE is facilitated through three learning pathways [31], aimed at utilizing the best available evidence at the point of care and conducting research or evaluating impact during routine care processes. The first pathway involves a connection between the clinical and research domains, facilitating intermittent information exchange. The second pathway provides an extended route for converting data into knowledge and incorporating evidence from EESs into networks of actors and organizations aimed at sharing and using (translated) knowledge at the point of care. The third pathway provides opportunities for learning from mirroring information and reflecting on the network functioning, patient experiences and individual (clinical) outcomes. Such external evidence and mirroring information are exchangeable through the data platform.

3.4.4 GOVERNANCE LAYER

Within this layer, governance within the SHE is identified to take place on different levels: at a global level, a decision-making level, an executive level, and an operative level. See Figure S6 in Additional File 4 for a visualizations of the model's overarching governance layer (using the intermediary analyses of '*governance'*). At each level, committees, boards, and/or assemblies are identified and play a role in overseeing the governance of the system. For example, at the decision-making level, an ethics and privacy board may be established to address issues related to privacy protection, protecting the interests of data-sharing parties, and offer advice regarding ethics and privacy for the network and its actors.

3.4.5 AGGREGATED MODEL

Figure 2 shows an aggregated SHE-model in which elements of all layers are merged. An extended visualization is shown in Figure S1 in Additional File 5. The aggregated model is static, depicting where data is generated and products are developed. However, Figure 3

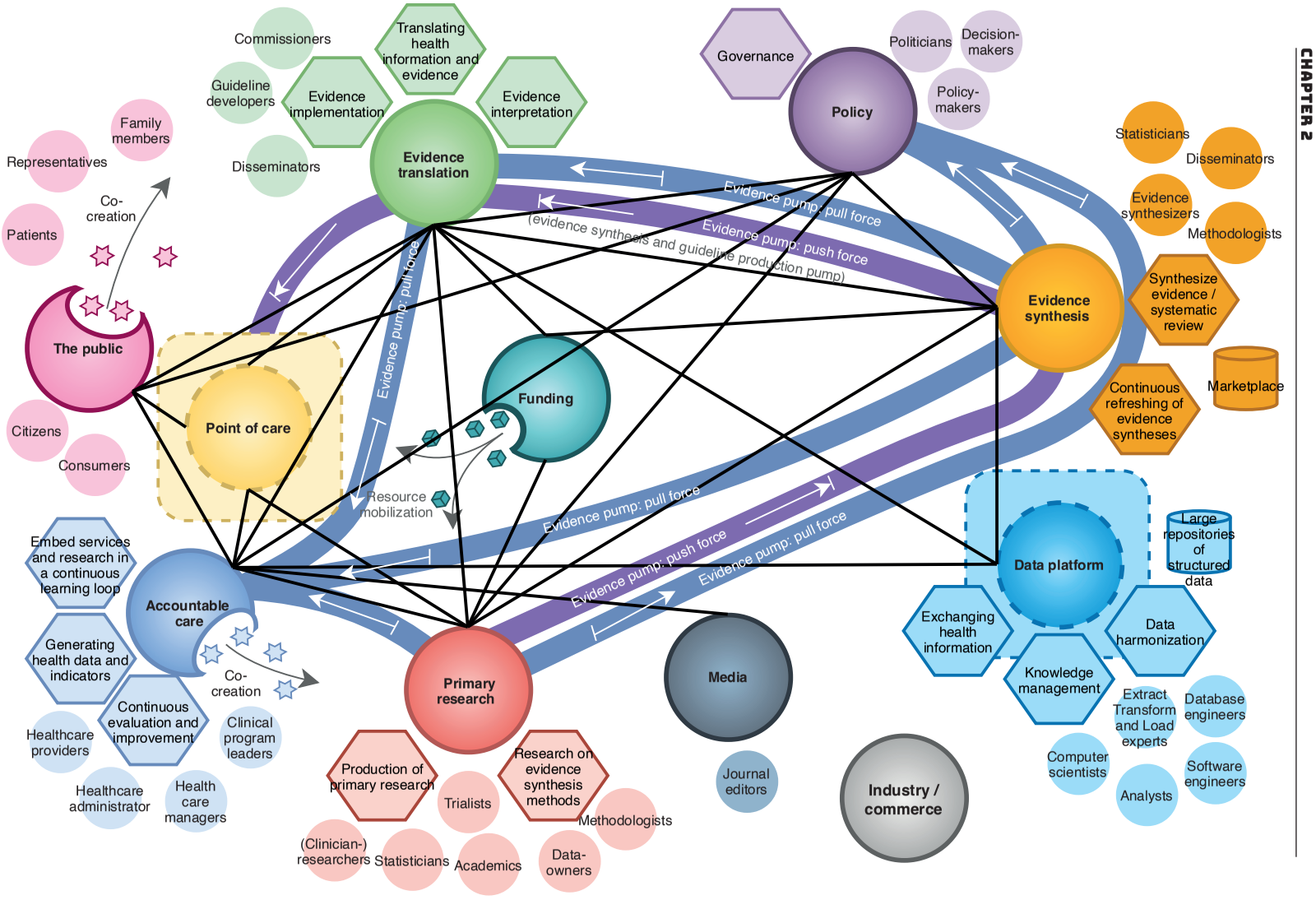

Commissioners
Translating health information and evidence
Guideline developers
Evidence implementation
Evidence interpretation
Disseminators
Evidence translation
Governance
Politicians
Decision-makers
Policy
Policy-makers
Family members
Representatives
Co-creation
Patients
The public
Statisticians
Disseminators
Evidence synthesizers
Methodologists
Evidence synthesis
Synthesize evidence / systematic review
Marketplace
Continuous refreshing of evidence syntheses
Evidence pump: pull force
Evidence pump: push force
(evidence syntheses and guideline production pump)
Evidence pump: pull force
Point of care
Funding
Citizens
Consumers
Resource mobilization
Evidence pump: pull force
Evidence pump: push force
Evidence pump: pull force
Embed services and research in a continuous learning loop
Accountable care
Generating health data and indicators
Co-creation
Continuous evaluation and improvement
Healthcare providers
Clinical program leaders
Healthcare administrator
Health care managers
Primary research
Production of primary research
Trialists
Research on evidence synthesis methods
Methodologists
(Clinician-) researchers
Statisticians
Academics
Data-owners
Media
Journal editors
Industry / commerce
Data platform
Large repositories of structured data
Exchanging health information
Data harmonization
Knowledge management
Extract Transform and Load experts
Database engineers
Computer scientists
Analysts
Software engineers

Figure 1. A basic representation of the foundation layer of a symbiotic healthcare ecosystem model. *The visualization shows an overview of links (black lines) between entities (large outlined colored circles) in the model. Actors (small colored circles) within entities, and the components (outlined hexagons) are added to show a possible foundation of a symbiotic healthcare model. This visualization was based on a more extended visualization (see Figure S1 in Additional File 3) and was simplified to increase the comprehensibility of the subsequent layers resulting in an aggregated model.*

represents a dynamic process showcasing the development of a CPG and a patient decision aid. These products are developed based on evidence syntheses derived from primary research, whereafter it is disseminated, applied, and evaluated in the point of care facilitating iterative learning cycles. In Figure 4, this dynamic process is mapped onto the aggregated SHE-model, illustrating the interplay between entities as data and products from one entity are used by other entities across the system. Key stakeholders, infrastructure, pathways, data, and products for this process can be identified this way. Throughout the system, co-creation ensures relevance and quality actions are aimed at ensuring trustworthiness of processes, procedures, and their resulting products. Moreover, observational data generated at the point of care may spark new hypotheses, prompting further primary research in controlled settings. This iterative cycle facilitates the initial testing of emerging hypotheses, driving innovation and advancement within the SHE.

4. DISCUSSION

The current study aimed to integrate the EES and LHS in a single SHE-model as a framework to speed up and improve complex processes from evidence generation to implementation and leaning in clinical practice. We discuss the essential aspects for connecting the EES and LHS effectively we identified from merging both systems. The SHE-framework (Figure 2) serves as a blueprint for connecting existing EESs with LHSs in the real world. It outlines various entities and actors contributing to generating and exchanging relevant data and products within the system through co-creation and quality actions. These data and evidence are produced for other entities in the system, both within and between sub-systems of a SHE. The blueprint supports the recent appeal to deliver pathways, infrastructure, enablers, and health benefits for the community[6] by connecting both systems. It might help in identifying relevant stakeholders, pathways, infrastructure, and products when complex and dynamic processes are mapped to the model overarching both the EES and LHS, similar to the example provided in Figure 4.

The LHS learning cycle integrates external evidence[8] and the EES contains the step "*evaluate and improve practice*". [5] It is crucial to understand the operational levels of both sub-systems. Here, the LHS is typically handling implementation and evaluation at a national or even a local level. [5] The EES undertakes evidence generation and synthetization at a global level. In this context, both the evidence translation and primary research entities seem to play a crucial role in bridging the LHS and EES. The evidence translation entity transfers global evidence from the EES for dissemination and implementation to the national or local LHS by contextualization. The primary research entity plays a role in both performing evaluations using real-world data in the LHS and generating primary evidence in the EES. The availability and accessibility of data and products plays a central role in the concept of an optimal interaction within and between systems, with a data platform in the system. This platform should be accessible by all participating entities. Although there are separate roles and responsibilities in the system, all can benefit from unhindered data exchange.

Data and evidence are generated throughout the SHE (see Figure S2 in S4 File). Real-

world data is generated from patient-clinician interaction in the LHS sub-system and studies are conducted during normal care. In the EES sub-system, the primary research entity conducts studies outside normal care to generate primary (experimental) data for syntheses by evidence synthesizers. These evidence syntheses may then be translated through the evidence translation entity (e.g. by developing CPGs, see Figures 3 and 4) and disseminated to and implemented in the LHS. Subsequently, data about evaluating the impact of translated products is generated at the point of care in the LHS. Products and data generated on one site in the system may be used to develop products at other sites in the SHE. As lots of data can be generated for lots of different purposes in a SHE, it becomes important to consider the nature of the generated data and products. For observational real-world data, for example, carefully designing and emulating a target trial may currently represent the best available method for causal inference[42] within the LHS for evaluative purposes. These evaluations have the potential to generate new hypotheses, which can subsequently be explored within the EES, leading to the development of new or updated contextualized products. Furthermore, observational data generated within the LHS can be used for other types of research within the EES, particularly where observational designs adequately provide answers to the research questions. Overall, there are various types of data and questions in a SHE. Different research and synthesis questions need different methods, data, and analytical approaches.[18]

Practice changing new insights and evidence should be integrated in CPGs as soon as possible. Thus, unnecessary delays should be prevented[2] in a SHE. One source of unnecessary delays could be the lack of relevant products in the EES. To mitigate this, co-creation initiatives throughout the SHE can ensure that products remain relevant for end-users within the system. Co-creation can aid in producing tailored outputs[4] of EES products for end-users. Another unnecessary delay in the system could be (not) finding relevant data and products, and not being able to exchange them efficiently within and between systems. Adding metadata to data making them findable, accessible, interoperable, reusable, and trustworthy[43] is therefore essential for a SHE. Together, the Fast Healthcare Interoperability Resources (FHIR) specification[44] and its extensions cover parts of both the LHS (with FHIR for clinical data) and the EES (with EBMonFHIR[44-46] and CPGonEBMonFHIR[47, 48] for primary research, systematic reviews, and CPGs). The first steps in connecting the EES and LHS using FHIR has been made, where CPG recommendation adherence was monitored using clinical data.[48] Furthermore, it crucial to share identified evidence gaps from systematic reviews and CPGs systematically with the primary research community, commissioners, and funders.

When exchanging data and products, users need to trust the quality. Additionally, there should be attention for privacy and security, especially when exchanging person-level data. All participating entities in a SHE could agree upon minimally acceptable practices for privacy, security, quality, and transparency. Much guidance for rigorous development and transparent reporting guidelines for products that would circulate the SHE already exists and are readily available. The primary research entity has reporting guidelines (e.g. CONSORT statement,[49] also see the EQUATOR network[50]) and even core outcome sets (e.g. see the COMET initiative[51] and ICHOM[52]). The evidence synthesizing entity has development handbooks for several types of reviews[53-58] and PRISMA reporting guidelines.[59] For guideline developers

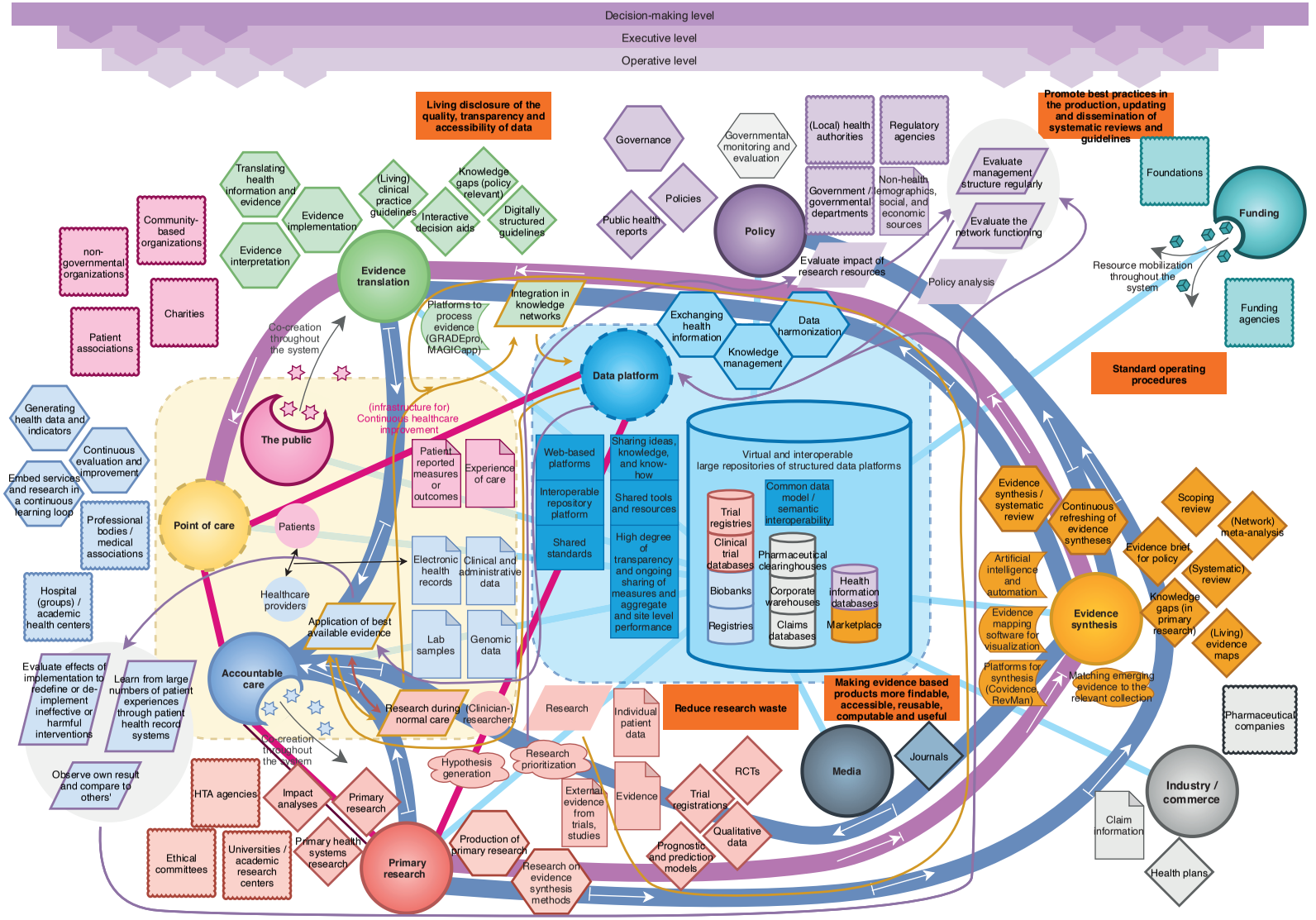

Decision-making level
Executive level
Operative level
Living disclosure of the quality, transparency and accessibility of data
Promote best practices in the production, updating and dissemination of systematic reviews and guidelines
Governance
Policies
Public health reports
Governmental monitoring and evaluation
Policy
(Local) health authorities
Regulatory agencies
Government / governmental departments
Non-health demographics, social, and economic sources
Evaluate management structure regularly
Evaluate the network functioning
Evaluate impact of research resources
Policy analysis
Foundations
Funding
Resource mobilization throughout the system
Funding agencies
Standard operating procedures
Translating health information and evidence
Evidence implementation
Evidence interpretation
(Living) clinical practice guidelines
Interactive decision aids
Knowledge gaps (policy relevant)
Digitally structured guidelines
Evidence translation
Community-based organizations
non-governmental organizations
Charities
Patient associations
Co-creation throughout the system
Platforms to process evidence (GRADEpro, MAGICapp)
Integration in knowledge networks
Exchanging health information
Data harmonization
Knowledge management
Data platform
Generating health data and indicators
Continuous evaluation and improvement
Embed services and research in a continuous learning loop
Professional bodies / medical associations
The public
(infrastructure for) Continuous healthcare improvement
Patient reported measures or outcomes
Experience of care
Point of care
Patients
Healthcare providers
Electronic health records
Clinical and administrative data
Lab samples
Genomic data
Application of best available evidence
Hospital (groups) / academic health centers
Web-based platforms
Interoperable repository platform
Shared standards
Sharing ideas, knowledge, and know-how
Shared tools and resources
High degree of transparency and ongoing sharing of measures and aggregate and site level performance
Virtual and interoperable large repositories of structured data platforms
Common data model / semantic interoperability
Trial registries
Clinical trial databases
Biobanks
Registries
Pharmaceutical clearinghouses
Corporate warehouses
Claims databases
Health information databases
Marketplace
Evidence synthesis / systematic review
Continuous refreshing of evidence syntheses
Scoping review
(Network) meta-analysis
Evidence brief for policy
(Systematic) review
Knowledge gaps (in primary research)
(Living) evidence maps
Artificial intelligence and automation
Evidence mapping software for visualization
Platforms for synthesis (Covidence, RevMan)
Evidence synthesis
Matching emerging evidence to the relevant collection
Evaluate effects of implementation to redefine or de-implement ineffective or harmful interventions
Learn from large numbers of patient experiences through patient health record systems
Accountable care
Observe own result and compare to others'
Co-creation throughout the system
Research during normal care
(Clinician-) researchers
Research
Individual patient data
Reduce research waste
Making evidence based products more findable, accessible, reusable, computable and useful
Media
Journals
Pharmaceutical companies
Industry / commerce
Claim information
Health plans
HTA agencies
Impact analyses
Primary research
Primary health systems research
Ethical committees
Universities / academic research centers
Primary research
Hypothesis generation
Research prioritization
Production of primary research
Research on evidence synthesis methods
External evidence from trials, studies
Evidence
Trial registrations
RCTs
Qualitative data
Prognostic and prediction models

Figure 2 . A visualization of a model for a symbiotic healthcare ecosystem. *The model displays the elements of the four model layers. It shows all the entities (solid outlined large circles) with their components (outlined hexagons), organizations (clouded squares) and actors (small circles, not outlined). Several entities generate data (outlined documents), develop products (outlined diamonds), have larger actions (parallelograms) and smaller actions (outlined clouds), and use tools (outlined curved rectangles). Data and information are stored in databases (outlined cylinders). The flow of evidence is presented as pink [push force] and blue [pull force] ribbons. Co-creation throughout the system is performed by the public and the accountable care entities (small stars), while the funding entity mobilizes resources throughout the system (small cubes). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. Entities have access to the data platform (blue lines) which has several features and characteristics (outlined blue rectangles). The infrastructure for continuous healthcare improvement is depicted as a network of thick magenta lines between the accountable care and primary research entities, together with the two locations in the system. Values for quality, evaluation, and learning (bright orange rectangles) will cause entities to perform quality actions (not shown) throughout the system. Three pathways for evaluation and learning processes are displayed as red (interaction between research and practice), yellow (data to knowledge, knowledge to performance, performance to data), and purple (interactions between stakeholders) arrows. Three levels of governance are shown (purple overarching beams).*

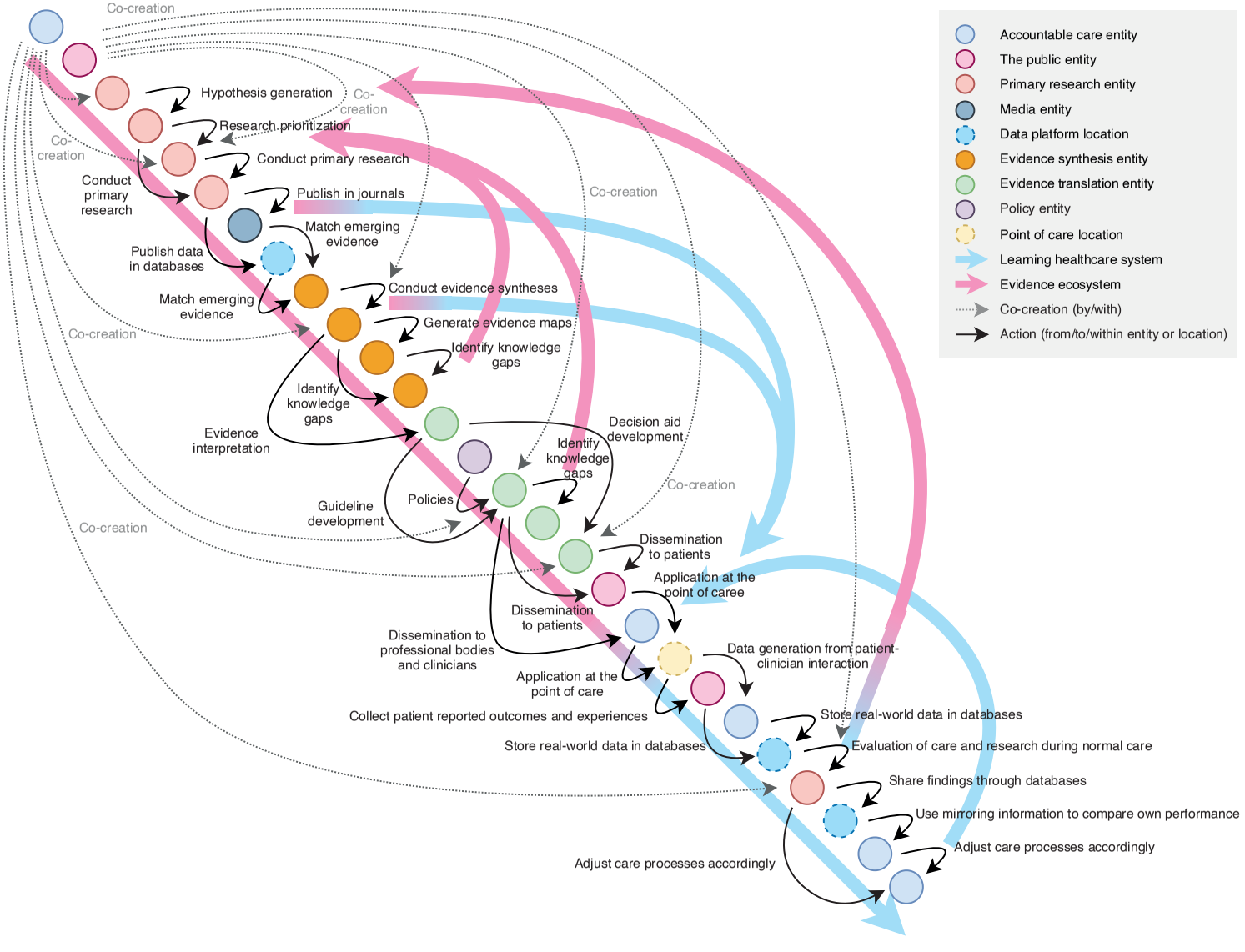

Co-creation
Hypothesis generation
Co-creation
Research prioritization
Co-creation
Conduct primary research
Conduct primary research
Publish in journals
Co-creation
Match emerging evidence
Publish data in databases
Conduct evidence syntheses
Match emerging evidence
Generate evidence maps
Co-creation
Identify knowledge gaps
Identify knowledge gaps
Evidence interpretation
Decision aid development
Identify knowledge gaps
Co-creation
Guideline development
Policies
Co-creation
Dissemination to patients
Application at the point of caree
Dissemination to patients
Dissemination to professional bodies and clinicians
Data generation from patient-clinician interaction
Application at the point of care
Collect patient reported outcomes and experiences
Store real-world data in databases
Store real-world data in databases
Evaluation of care and research during normal care
Share findings through databases
Use mirroring information to compare own performance
Adjust care processes accordingly
Adjust care processes accordingly
Accountable care entity
The public entity
Primary research entity
Media entity
Data platform location
Evidence synthesis entity
Evidence translation entity
Policy entity
Point of care location
Learning healthcare system
Evidence ecosystem
Co-creation (by/with)
Action (from/to/within entity or location)

Figure 3. A process chain of evidence generation, evidence synthesis, evidence translation, application, evaluation, and learning. *The figure visualizes a pathway from the top left to the bottom right showing the involved symbiotic healthcare ecosystem entities (outlined circles) and locations (dashed circles) in a process from evidence generation to application, whereafter provided care is evaluated and learned from. Black arrows show actions undertaken by entities. Action may be performed within an entity (actions connect circles of the same color) or from/to different entities or locations (actions connect different colored circles). Grey dashed arrows indicate where the accountable care entity (e.g. healthcare providers) and the public entity (e.g. patients, representatives) might co-create in the process. The thick pink and blue arrows show where the evidence ecosystem and learning healthcare sub-systems respectively can operate in the background of the process.*

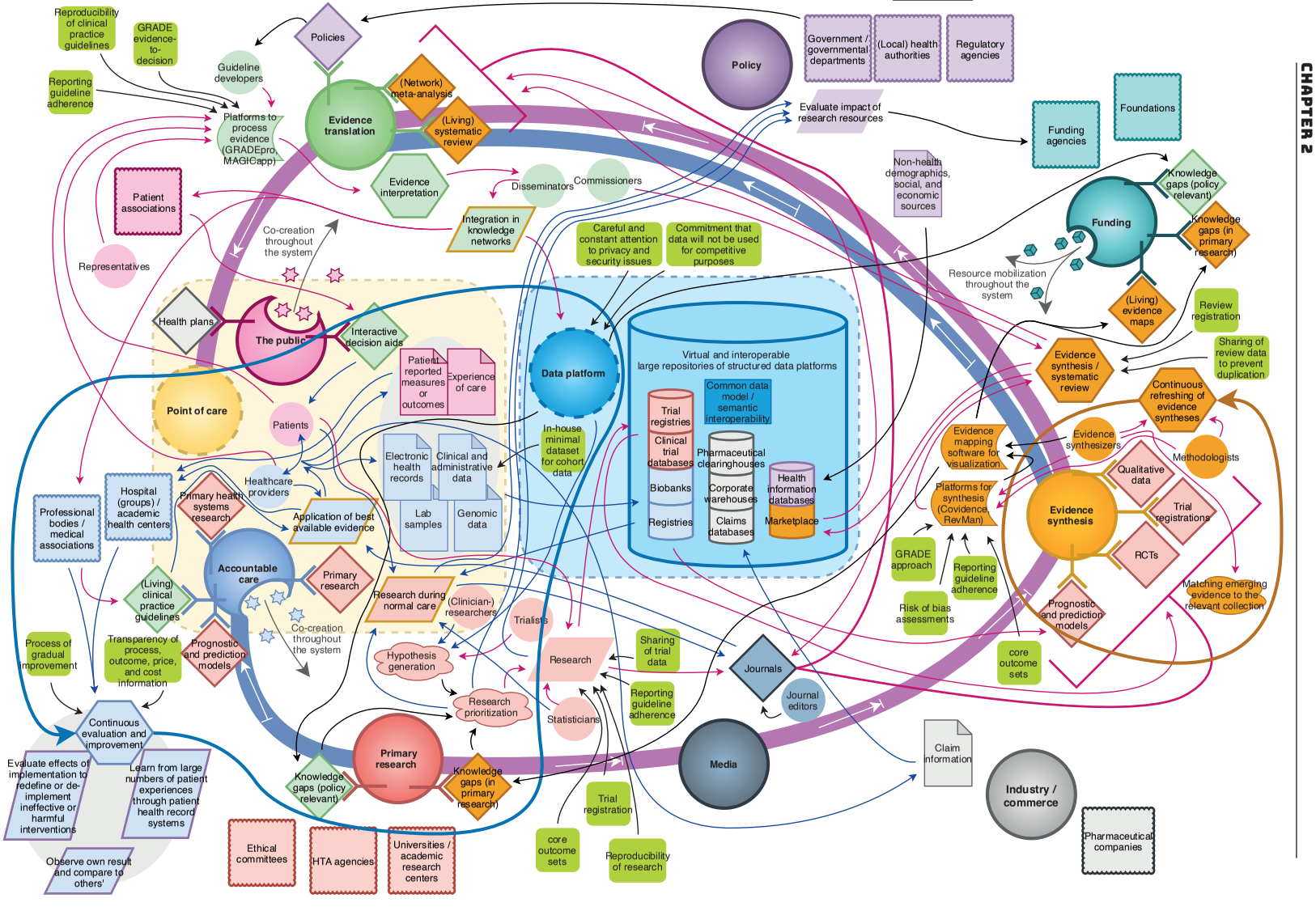

Reproducibility of clinical practice guidelines
GRADE evidence-to-decision
Reporting guideline adherence
Guideline developers
Policies
Platforms to process evidence (GRADEpro, MAGICapp)
Evidence translation
(Network) meta-analysis
(Living) systematic review
Evidence interpretation
Disseminators
Commissioners
Integration in knowledge networks
Patient associations
Representatives
Co-creation throughout the system
Policy
Government / governmental departments
(Local) health authorities
Regulatory agencies
Evaluate impact of research resources
Funding agencies
Foundations
Non-health demographics, social, and economic sources
Funding
Knowledge gaps (policy relevant)
Knowledge gaps (in primary research)
Resource mobilization throughout the system
(Living) evidence maps
Review registration
Careful and constant attention to privacy and security issues
Commitment that data will not be used for competitive purposes
Health plans
The public
Interactive decision aids
Point of care
Patients
Patient reported measures or outcomes
Experience of care
Data platform
Virtual and interoperable large repositories of structured data platforms
Common data model / semantic interoperability
Trial registries
Clinical trial databases
Biobanks
Registries
Pharmaceutical clearinghouses
Corporate warehouses
Claims databases
Health information databases
Marketplace
In-house minimal dataset for cohort data
Electronic health records
Clinical and administrative data
Lab samples
Genomic data
Healthcare providers
Application of best available evidence
Primary health systems research
Hospital (groups) / academic health centers
Professional bodies / medical associations
Accountable care
Primary research
(Living) clinical practice guidelines
Research during normal care
(Clinician-) researchers
Trialists
Co-creation throughout the system
Process of gradual improvement
Transparency of process, outcome, price, and cost information
Prognostic and prediction models
Hypothesis generation
Research
Research prioritization
Statisticians
Sharing of trial data
Reporting guideline adherence
Journals
Journal editors
Media
Claim information
Industry / commerce
Pharmaceutical companies
Evidence synthesis / systematic review
Sharing of review data to prevent duplication
Continuous refreshing of evidence syntheses
Evidence synthesizers
Methodologists
Evidence mapping software for visualization
Platforms for synthesis (Covidence, RevMan)
Evidence synthesis
Qualitative data
Trial registrations
RCTs
Prognostic and prediction models
GRADE approach
Reporting guideline adherence
Risk of bias assessments
core outcome sets
Matching emerging evidence to the relevant collection
Continuous evaluation and improvement
Evaluate effects of implementation to redefine or de-implement ineffective or harmful interventions
Learn from large numbers of patient experiences through patient health record systems
Observe own result and compare to others'
Primary research
Knowledge gaps (policy relevant)
Knowledge gaps (in primary research)
Ethical committees
HTA agencies
Universities / academic research centers
core outcome sets
Trial registration
Reproducibility of research

Figure 4. A visualization of a symbiotic ecosystem in process: guideline development, dissemination, evaluation, and learning. *The figure shows the entities in the system developing a clinical practice guideline and learning from the application of best evidence through the guideline. The pink (push force) and blue (pull force) ribbons show the flow of evidence through the system's evidence pumps. The thin red arrows show the network of primary evidence production, evidence synthesis, and evidence translation towards entities in the point of care. The thick orange arrow depicts the continuous nature of refreshing or updating the evidence syntheses. Identified evidence gaps by the evidence synthesis entity may be used by the primary research and funding entities to prioritize research. The thin blue arrows show the network of cyclical evaluation and learning, where the application of guidelines during the patient-clinician interaction and within hospitals can be evaluated and learned from. The thick blue arrow depicts continuous evaluation and improvement in healthcare provision. Learning can also form new hypotheses that need to be tested in the primary research entity. Furthermore, the thin black arrows are supporting processes in the network where quality actions (bright green fields) are in place to ensure trustworthy processes and products throughout the system. In a functioning symbiotic healthcare ecosystem, the developed products (outlined diamonds) by one entity are used elsewhere in the system.*

there are essential domains, [60] handbooks, [61, 62] a framework for adaptations, [63] guidance for assessing the certainty of a body of evidence[64] (along with numerous ongoing GRADE publications in peer-reviewed journals), and even a framework to transparently report the considerations from evidence to recommendations. [65] However, it should be acknowledged that products in the EES take time for rigorous development. Speeding up the process might compromise the quality[66] and building a body of evidence through replication obviously takes time. Such delay is necessary to gain certainty about the safety and efficacy of new developments[2] and may therefore be a necessity in a SHE. This should not mean that there is no opportunity to rapidly learn in the local LHS. Evaluating their practices and patient experiences while mirroring processes and procedures from other practices in the system might create quality improvement through learning from these evaluations and reflections. This, however, might be more on the level of improving internal processes, diagnostics and prognosis, teamwork, culture, knowledge exchange, implementation, resource expenditure, (patient) experiences, learning opportunities, and technical skills, etc. instead of introducing new treatments with promising results from observational data to improve health outcomes in a practice.

Factors influencing the success or failure of the system may be accepting of the model itself[12] and the willingness to participate[11] or support.[24] Co-creating significant organizational changes[22] may be required as community-building and the agility to adapt to changing needs[22] seem important. Furthermore, it might be important to reach agreement on statues and financial contributions[11] (including investing dedicated resources[7]), understand individual perspectives and assumptions, [20] and match expectations and activities[11] within the system. Suitable personnel[11] with expertise[7] having incentives[23] and appropriate rewards and recognition[12] may be important in a system with a precompetitive spirit[20] and without competing initiatives. [11]

It is essential that organizations in the SHE feel responsible to deliver and support the SHE, where they focus on exchanging and maintaining their data and products sustainably. Organizations should initiate changes to address such responsibilities adequately, anticipating that new developments could evolve their internal processes over time. For example, organizations may need to adapt their current processes to enable continuous maintenance of their products by setting-up internal processes for living systematic reviews, [67] living CPGs, [68] or tailored maintenance strategies[69] to reduce unnecessary delays caused by outdated updating procedures (e.g. update every 3 years). When challenges are overcome, organizations should adjust their internal processes that now create unnecessary delays in the pathway from evidence generation to implementation. These processes may need revisitations over time, as new developments could create new unnecessary delays in the pathway.

One of the limitations in this study is the discrepancy in the search strategy for EES and LHS. While we focused on reviews describing the LHS, we did not limit the search for EES characteristics to reviews. This approach, motivated by practical reasons, may have resulted in overlooking primary studies that describe LHS characteristics not captured in reviews and could furthermore have limited our discussion of factors influencing the success of the system. Nonetheless, we believe a wide variety of characteristics were captured making it able to develop a SHE-model. Another limitation is that the visualizations of the SHE might

show layers and models which reflect our own interpretations and perspectives. This could deviate from how such a model would operate in the real world. In fact, the visualizations altogether are envisioning a system and its functioning that currently solely seems theoretical and forward-looking. However, the presented SHE-model might clarify how complex processes in such systems could function.

5. CONCLUSION

To capitalize on a symbiotic system focused on optimal evidence generation and implementation, it seems important to understand important aspects for connecting the EES and LHS sub-systems (see Box 1 for a summary). Unhindered availability of relevant and trustworthy data and products in relevant formats seems to be a key factor in this symbiotic system. The nature, relevance, and use of data generated throughout the system should be understood and system-wide agreements on privacy, security, transparency, quality, and data exchange are needed for trust. Co-creation and sharing knowledge gaps could furthermore ensure the relevance of products in the SHE. The SHE-framework can be used as a blueprint to map complex processes over its connected sub-systems. This could identify relevant stakeholders, pathways, infrastructure, data, and products for existing or new real-world processes across these sub-systems. Considering such complex processes through the SHE-framework's lens supports identifying (un)necessary delays in the identified pathways (including organizational processes) from evidence generation to implementation and learning. Stepwise implementation of a symbiotic healthcare system has the potential to speed up and improve evidence generation, implementation and learning in clinical practice.

Box 1. Discussed aspects for a symbiotic healthcare ecosystem focusing on an optimal pathway from evidence generation to implementation and learning.

Levels

- An understanding that both sub-systems operate on different levels: the EES operates on a global level, where the LHS operates on the national or local level. This includes:
  - » An understanding that evidence translators contextualize global evidence for the national or local context in the LHS.
  - » An understanding that the primary research entity may play a role in both national/local evaluations in the LHS and in the generation of primary evidence used globally in the EES.

Relevance of data and products in the system

- An understanding of the nature, relevance, and use of data and products generated across the system.
- The inclusion of clinicians and patients in co-creating products throughout the entire system.
- Usage of relevant formats of data and products for other entities and actors in the system.

Infrastructure and exchangeability

- Usage of linked, interoperable, and searchable databases to efficiently transfer data and products between and within sub-systems:
  - » Making data and products in databases searchable for all participating entities by adding meta-data.
  - » Usage of a common language (or languages that are compatible) to communicate when exchanging data.

Trustworthiness

- Agreements on the minimal acceptable practices to:
  - » Ensure privacy and security.
  - » Generate and develop trustworthy data and products in the EES system.
  - » Evaluate (contextualized) impact and effectiveness claims from the EES in the LHS.
  - » Transparently report products and data.
- An understanding of how trustworthy products may take time to be developed in the EES before implemented in the LHS with an acceptable certainty.

Roles and responsibilities

- Agreements on the minimal acceptable practices in the system may need revisitation over time to remove unnecessary delays caused by outdated internal organizational processes.
- An understanding that participating organizations remain responsible for the currency and rigor of their circulating products in the system, including:
  - » An understanding that participating organizations may need to adapt their internal processes to meet such responsibilities.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

All author and non-author contributions are made transparent using the structured Contributor Role Taxonomy (CRediT) described at https://doi.org/10.1002/leap.1210.

MSO: Conceptualization, methodology, investigation, formal analysis, data curation, writing – original draft, visualization
FMJ: Validation, data curation, writing – review & editing
MJvdL: Conceptualization, methodology, supervision, writing – review & editing
LH: Conceptualization, methodology, supervision, writing – review & editing

DATA AVAILABILITY STATEMENT

All relevant data are in the manuscript and its Additional files.

LITERATURE LIST

1. Institute of Medicine, Roundtable on Evidence-Based Medicine. In: Olsen L, Aisner D, McGinnis JM, editors. The Learning Healthcare System: Workshop Summary. Washington (DC): National Academies Press (US)Copyright © 2007, National Academy of Sciences.; 2007.
2. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med. 2011;104(12):510-20.
3. Hanney SR, Castle-Clarke S, Grant J, Guthrie S, Henshall C, Mestre-Ferrandiz J, et al. How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice. Health Res Policy Syst. 2015;13:1.
4. 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. J Clin Epidemiol. 2020;123:153-61.
5. Vandvik PO, Brandt L. Future of Evidence Ecosystem Series: Evidence ecosystems and learning health systems: why bother? J Clin Epidemiol. 2020;123:166-70.
6. Rajit D, Johnson A, Callander E, Teede H, Enticott J. Learning health systems and evidence ecosystems: a perspective on the future of evidence-based medicine and evidence-based guideline development. Health Res Policy Syst. 2024;22(1):4.
7. Easterling D, Perry AC, Woodside R, Patel T, Gesell SB. Clarifying the concept of a learning health system for healthcare delivery organizations: Implications from a qualitative analysis of the scientific literature. Learn Health Syst. 2022;6(2):e10287.
8. Guise JM, Savitz LA, Friedman CP. Mind the Gap: Putting Evidence into Practice in the Era of Learning Health Systems. J Gen Intern Med. 2018;33(12):2237-9.
9. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, Contreras JAP, Borrero AMP, Patiño-Lugo DF, Valencia DAC. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst. 2020;18(1):74.
10. Akl EA, Haddaway NR, Rada G, Lotfi T. Future of Evidence Ecosystem Series: Evidence synthesis 2.0: when systematic, scoping, rapid, living, and overviews of reviews come together. J Clin Epidemiol. 2020;123:162-5.
11. Bogaert P, van Oers H, Van Oyen H. Towards a sustainable EU health information system infrastructure: A consensus driven approach. Health Policy. 2018;122(12):1340-7.
12. Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. Ann Intern Med. 2020;173(12):1015-7.
13. Cartabellotta A, Tilson JK. The ecosystem of evidence cannot thrive without efficiency of knowledge generation, synthesis, and translation. J Clin Epidemiol. 2019;110:90-5.
14. 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. J Clin Epidemiol. 2020;123:143-52.

15. Dumitrescu A, Granados A, Wallace J, Watson S. Demand-driven evidence network in Europe. Bull World Health Organ. 2006;84(1):2.
16. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, Gruen RL. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11(2):e1001603.
17. Funk T, Sharma T, Chapman E, Kuchenmüller T. Translating health information into policy-making: A pragmatic framework. Health Policy. 2022;126(1):16-23.
18. Gough D, Thomas J, Oliver S. Clarifying differences between reviews within evidence ecosystems. Syst Rev. 2019;8(1):170.
19. Jordan Z, Lockwood C, Aromataris E, Pilla B, Porritt K, Klugar M, et al. JBI series paper 1: Introducing JBI and the JBI Model of EHBC. J Clin Epidemiol. 2022;150:191-5.
20. Lovestone S. The European medical information framework: A novel ecosystem for sharing healthcare data across Europe. Learn Health Syst. 2020;4(2):e10214.
21. Mathews C, Goga A, Loveday M, Zembe W, Daviaud E, Siegfried N, Lewin S. Moving towards universal health coverage: Strengthening the evidence ecosystem for the South African health system. S Afr Med J. 2019;109(11b):8-14.
22. Pilla B, Jordan Z, Christian R, Kynoch K, McInerney P, Cooper K, et al. JBI series paper 4: the role of collaborative evidence networks in promoting and supporting evidence-based health care globally: reflections from 25 years across 38 countries. J Clin Epidemiol. 2022;150:210-5.
23. Shepherd JP. How to achieve more effective services: the evidence ecosystem. 2014.
24. Stewart R, Dayal H, Langer L, Van Rooyen C. The evidence ecosystem in South Africa: growing resilience and institutionalisation of evidence use. Palgrave Communications. 2019;5(1).
25. Apfeld JC, Deans KJ. Learning health systems and the future of clinical research. J Pediatr Surg. 2020;55s:51-3.
26. Atkins D, Kilbourne AM, Shulkin D. Moving From Discovery to System-Wide Change: The Role of Research in a Learning Health Care System: Experience from Three Decades of Health Systems Research in the Veterans Health Administration. Annu Rev Public Health. 2017;38:467-87.
27. Britto MT, Fuller SC, Kaplan HC, Kotagal U, Lannon C, Margolis PA, et al. Using a network organisational architecture to support the development of Learning Healthcare Systems. BMJ Qual Saf. 2018;27(11):937-46.
28. Budrionis A, Bellika JG. The Learning Healthcare System: Where are we now? A systematic review. J Biomed Inform. 2016;64:87-92.
29. Casey JD, Courtright KR, Rice TW, Semler MW. What can a learning healthcare system teach us about improving outcomes? Curr Opin Crit Care. 2021;27(5):527-36.
30. Committee on the Learning Health Care System in America, Institute of Medicine. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
31. de Bruin J, Bos C, Struijs JN, Drewes HW, Baan CA. Conceptualizing learning

health systems: A mapping review. Learn Health Syst. 2023;7(1):e10311.
32. Deans KJ, Sabihi S, Forrest CB. Learning health systems. Semin Pediatr Surg. 2018;27(6):375-8.
33. Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26(2):w107-18.
34. Institute of Medicine. In: Grossmann C, Powers B, McGinnis JM, editors. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
35. Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. In: Olsen L, Saunders RS, McGinnis JM, editors. Patients Charting the Course: Citizen Engagement and the Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
36. Leung TI, Dumontier M. FAIR Principles for Clinical Practice Guidelines in a Learning Health System. Stud Health Technol Inform. 2019;264:1690-1.
37. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Digit Med. 2021;4(1):2.
38. Platt JE, Raj M, Wienroth M. An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review. J Med Internet Res. 2020;22(3):e17026.
39. Roundtable on Translating Genomic-Based Research for Health, Board on Health Sciences Policy, Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. Genomics-Enabled Learning Health Care Systems: Gathering and Using Genomic Information to Improve Patient Care and Research: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2015 by the National Academy of Sciences. All rights reserved.; 2015.
40. Roundtable on Value and Science-Driven Health Care, Forum on Drug Discovery Development and Translation, Board on Health Sciences Policy, Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. Large Simple Trials and Knowledge Generation in a Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
41. A Learning Health System Activity, Roundtable on Value Science-Driven Health Care, Institute of Medicine. Observational Studies in a Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
42. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
43. Alper BS, Flynn A, Bray BE, Conte ML, Eldredge C, Gold S, et al. Categorizing metadata to help mobilize computable biomedical knowledge. Learn Health Syst. 2022;6(1):e10271.
44. HL7 International. FHIR Overview. 2023. Available from: https://www.hl7.org/fhir/overview.html
45. Alper BS, Dehnbostel J, Afzal M, Subbian V, Soares A, Kunnamo I, et al. Making

science computable: Developing code systems for statistics, study design, and risk of bias. J Biomed Inform. 2021;115:103685.
46. HL7 International. EBMonFHIR. 2023.
47. Lichtner G, Alper BS, Jurth C, Spies C, Boeker M, Meerpohl JJ, von Dincklage F. Representation of evidence-based clinical practice guideline recommendations on FHIR. J Biomed Inform. 2023;139:104305.
48. Lichtner G, Spies C, Jurth C, Bienert T, Mueller A, Kumpf O, et al. Automated Monitoring of Adherence to Evidenced-Based Clinical Guideline Recommendations: Design and Implementation Study. J Med Internet Res. 2023;25:e41177.
49. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Bmj. 2010;340:c332.
50. EQUATOR Network. EQUATOR Network: Enhancing the QUAlity and Transparency of health Research. 2023. Available from: https://www.equator-network.org/
51. COMET Initiative. COMET Initiative: Core Outcome Measures in Effectiveness Trials. 2023. Available from: https://www.comet-initiative.org/, Accessed on 25 July 2023
52. International Consortium for Health Outcomes Measurement. Patient-Centered Outcome Measures: Start measuring what matters most to patients. 2023. Available from: https://www.ichom.org/patient-centered-outcome-measures/
53. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane; 2022. Available from: www.training.cochrane.org/handbook
54. Deeks JJ, Bossuyt PM, Leeflang M, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy version 2.0 (updated July 2023). Cochrane; 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current
55. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021;130:13-22.
56. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
57. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.
58. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.
59. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372:n71.
60. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health

care. Cmaj. 2010;182(18):E839-42.
61. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. 2023. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction
62. Scottish Intercollegiate Guidelines Network. A guideline developer's handbook (SIGN publication no. 50). Edinburgh: SIGN; 2019. Available from: http://www.sign.ac.uk
63. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L, et al. Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. Int J Qual Health Care. 2006;18(3):167-76.
64. Schünemann H, Brożek J, Guyatt G, Oxmann A. GRADE Handbook. 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html [Accessed on: 9 August 2023]
65. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. Bmj. 2016;353:i2089.
66. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. Bmj. 2016;354:i5191.
67. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol. 2017;91:47-53.
68. El Mikati IK, Khabsa J, Harb T, Khamis M, Agarwal A, Pardo-Hernandez H, et al. A Framework for the Development of Living Practice Guidelines in Health Care. Ann Intern Med. 2022;175(8):1154-60.
69. Oerbekke MS, Elbers RG, van der Laan MJ, Hooft L. Designing tailored maintenance strategies for systematic reviews and clinical practice guidelines using the Portfolio Maintenance by Test-Treatment (POMBYTT) framework. BMC Med Res Methodol. 2024;24(1):29.

CHAPTER 2
Additional files

ADDITIONAL FILE 1

EVIDENCE ECOSYSTEMS

Search string (MEDLINE via PubMed)
"evidence ecosystem"[tiab] OR "evidence ecosystems"[tiab] OR "evidence eco-system"[-tiab] OR "evidence eco-systems"[tiab] OR "evidence universe"[tiab] OR "evidence network"[tiab] OR "evidence networks"[tiab]

Inclusion criterium
- When any characteristic of (parts of) an EES or a similar concept in a health care context was described in a report or study

Exclusion criteria
- When describing living systematic review or living guideline methodology when not placed in the context of an EES
- When solely describing results or methodologies for systematic reviews and/or (network) meta-analyses
- Conference abstracts

LEARNING HEALTHCARE SYSTEMS

Search string (MEDLINE via PubMed)
("learning health"[ti] OR "learning healthcare"[ti]) AND ("Systematic Review"[pt] OR "Review"[pt])

Inclusion criterium
- When any characteristic or (parts of) an LHS was described in a (systematic) review

Exclusion criteria
- When not having LHS as a main topic
- When solely reporting (clinical) results and/or compared outcomes in an LHS,
- When describing the management of a health condition within an LHS
- When describing a system specifically developed for one or multiple clinical conditions, health professions or clinical fields,
- When describing a different system in analogy to an LHS
- When describing a process to promote, enable, or move towards an LHS
- Conference abstracts

ADDITIONAL FILE 2

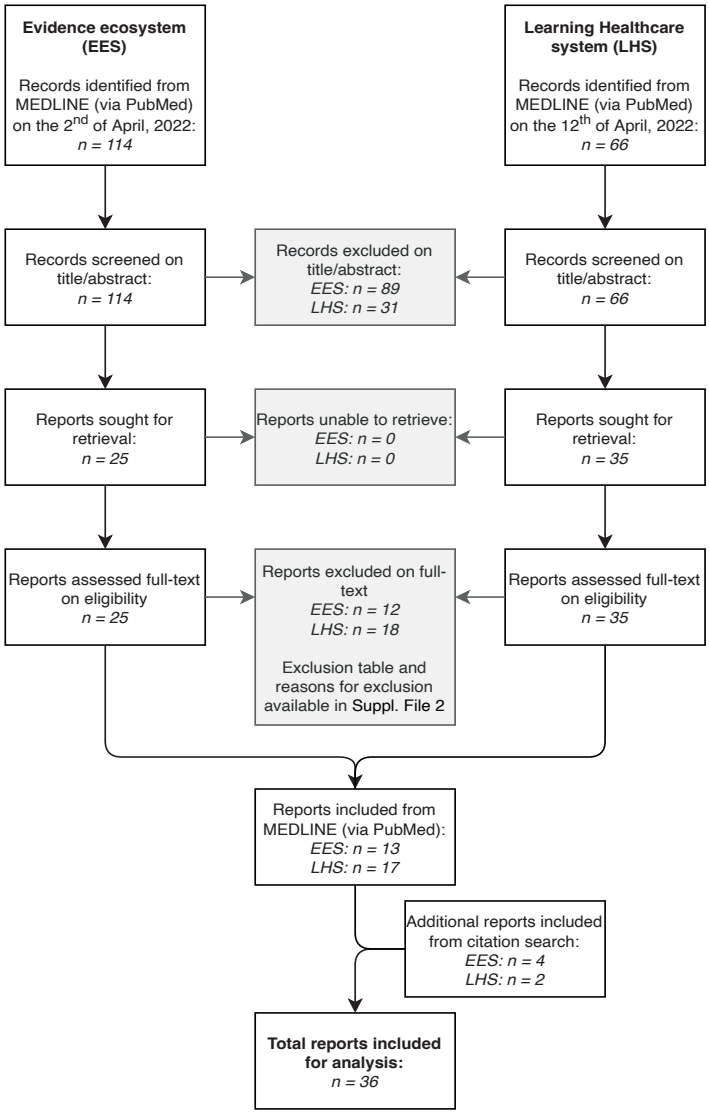


Figure S1. Flow of study selection. *A flow diagram describing the study selection process.*

Table S1 - Exclusion table

Reference	Reason for exclusion
Evidence ecosystems	
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. J Clin Epidemiol. 2020 Jul;123:135-142. doi: 10.1016/j.jclinepi.2020.01.024. Epub 2020 Mar 4. PMID: 32145367.	Seems to concern evidence synthesis system rather than evidence ecosystem itself, like a problem-statement paper / introduction to evidence synthesis ecosystems.
Cadarette SM, He N, Chaudhry M, Dolovich L. The Ontario Pharmacy Evidence Network Interactive Atlas of Professional Pharmacist Services. Can Pharm J (Ott). 2021 May 28;154(3):153-159. doi: 10.1177/17151635211004969. PMID: 34104268; PMCID: PMC8165887.	Does not seem to be in the context of evidence ecosystems.
Chaudhry M, He N, Waite NM, Houle SKD, Kwong JC, Cadarette SM. The Ontario Pharmacy Evidence Network Atlas of Community Pharmacy Influenza Immunizations. Can Pharm J (Ott). 2021 Aug 11;154(5):305-311. doi: 10.1177/17151635211034207. PMID: 34484480; PMCID: PMC8408905.	Does not seem to be in the context of evidence ecosystems.
Cooke M, Waite N, Cook K, Milne E, Chang F, McCarthy L, Sproule B. Incorporating sex, gender and vulnerable populations in a large multisite health research programme: The Ontario Pharmacy Evidence Network as a case study. Health Res Policy Syst. 2017 Mar 20;15(1):20. doi: 10.1186/s12961-017-0182-z. PMID: 28320403; PMCID: PMC5360067.	Does not seem to be in the context of evidence ecosystems.
Karlsson LE, Takahashi R. A Resource for Developing an Evidence Synthesis Report for Policy-Making [Internet]. Copenhagen: WHO Regional Office for Europe; 2017. PMID: 28956894.	Seems to be evidence synthesis methodology to inform policy-decisions; Describes steps for evidence synthesis, does not seem to describe characteristics of an evidence ecosystem per se.
Lewin S, Glenton C. Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the World Health Organization. Int J Equity Health. 2018 Sep 24;17(1):126. doi: 10.1186/s12939-018-0841-x. PMID: 30244675; PMCID: PMC6151925.	Seems to mainly focus on new developments in qualitative research rather than describing characteristics of an evidence ecosystem: Seems to mainly describe how qualitative research ended up in evidence synthesis which ended up in policy/guidelines; does not seem to mainly describe an evidence ecosystem or its characteristics in itself.
Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation-challenges-success factors. Onkologe (Berl). 2018;24(Suppl 2):91-98. doi: 10.1007/s00761-018-0358-3. Epub 2018 Jun 7. PMID: 30464373; PMCID: PMC6224010.	Seems to focus on real-world evidence in research rather than describing characteristics of an evidence ecosystem.
etzendorf MI, Featherstone RM. Evaluation of the comprehensiveness, accuracy and currency of the Cochrane COVID-19 Study Register for supporting rapid evidence synthesis production.	Seems to describe methodology to support rapid/living evidence synthesis, comprehensiveness of database.

MRes Synth Methods. 2021 Sep;12(5):607-617. doi: 10.1002/jrsm.1501. Epub 2021 Aug 1. PMID: 34089295; PMCID: PMC8242693.	
Pan J, Zhong Y, Young S, Niezink NMD. Collaboration on evidence synthesis in Africa: a network study of growing research capacity. Health Res Policy Syst. 2021 Sep 19;19(1):126. doi: 10.1186/s12961-021-00774-2. PMID: 34538255; PMCID: PMC8451124.	Seems to be an evaluation of collaborations in evidence synthesis.
Piccini JP, Kong DF. Mixed treatment comparisons for atrial fibrillation: evidence network or bewildering entanglement? Europace. 2011 Mar;13(3):295-6. doi: 10.1093/europace/eur029. PMID: 21345924.	Seems to be an editorial discussing clinical results and (network) meta-analyses.
Rowlands G, Trezona A, Russell S, Lopatina M, Pelikan J, Paasche-Orlow M, Drapkina O, Kontsevaya A, Sørensen K. What is the evidence on the methods, frameworks and indicators used to evaluate health literacy policies, programmes and interventions at the regional, national and organizational levels? [Internet]. Copenhagen: WHO Regional Office for Europe; 2019. PMID: 31693320.	Seems to be a scoping review for indicators on health literacy.
Wang Q, Li N, Li J, He Y, Li Y, Zhong D, Liu X, Fan J, Jin R, Kang D, Zhang Y. A Protocol of a Guideline to Establish the Evidence Ecosystem of Acupuncture. Front Med (Lausanne). 2022 Feb 15;8:711197. doi: 10.3389/fmed.2021.711197. PMID: 35252220; PMCID: PMC8896352.	Seems to be a protocol for a guideline, does not seem to describe ecosystem characteristics.
Learning Healthcare Systems	
Cahan A, Cimino JJ. A Learning Health Care System Using Computer-Aided Diagnosis. J Med Internet Res. 2017 Mar 8;19(3):e54. doi: 10.2196/jmir.6663. PMID: 28274905; PMCID: PMC5362695.	Seems to concern 'decision support system' as a road to/ enabler of a learning healtcare system, not necessarily a LHS characteristic.
Ellis LA, Sarkies M, Churruca K, Dammery G, Meulenbroeks I, Smith CL, Pomare C, Mahmoud Z, Zurynski Y, Braithwaite J. The Science of Learning Health Systems: Scoping Review of Empirical Research. JMIR Med Inform. 2022 Feb 23;10(2):e34907. doi: 10.2196/34907. Erratum in: JMIR Med Inform. 2022 Aug 4;10(8):e41424. PMID: 35195529; PMCID: PMC8908194.	Seems to concern: emperical contributions, areas of research, study design and methodology, implementation outcomes and determinants, and implementation frameworks and tools.
Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: a systematic review. BMC Health Serv Res. 2021 Mar 5;21(1):200. doi: 10.1186/s12913-021-06215-8. PMID: 33663508; PMCID: PMC7932903.	Seems to describe outcomes of LHS and benefits to clinicians/patients/organizations/system-level, does not seem to describe characteristics of an LHS.
Fiscella K, Tobin JN, Carroll JK, He H, Ogedegbe G. Ethical oversight in quality improvement and quality improvement research: new approaches to promote a learning health care system. BMC Med Ethics. 2015 Sep 17;16(1):63. doi: 10.1186/s12910-015-0056-2. PMID: 26383770;	Seems to discuss quality improvement and quality improvement research in light of ethics.

PMCID: PMC4574354.	
Ginsburg GS, Kuderer NM. Comparative effectiveness research, genomics-enabled personalized medicine, and rapid learning health care: a common bond. J Clin Oncol. 2012 Dec 1;30(34):4233-42. doi: 10.1200/JCO.2012.42.6114. Epub 2012 Oct 15. PMID: 23071236; PMCID: PMC3504328.	Seems to mainly focus on comparative effectiveness research and personalized medicine while not primarily having learning health systems as a topic.
Institute of Medicine (US) Roundtable on Evidence-Based Medicine. The Learning Healthcare System: Workshop Summary. Olsen L, Aisner D, McGinnis JM, editors. Washington (DC): National Academies Press (US); 2007. PMID: 21452449.	Seems to describe problems, efforts, and alternatives while building an argument for an LHS.
Institute of Medicine (US) and National Academy of Engineering (US) Roundtable on Value & Science-Driven Health Care. Engineering a Learning Healthcare System: A Look at the Future: Workshop Summary. Washington (DC): National Academies Press (US); 2011. PMID: 21977540.	Seems to discuss concepts from engineering as an opportunity to establish an LHS.
Lim HC, Austin JA, van der Vegt AH, Rahimi AK, Canfell OJ, Mifsud J, Pole JD, Barras MA, Hodgson T, Shrapnel S, Sullivan CM. Toward a Learning Health Care System: A Systematic Review and Evidence-Based Conceptual Framework for Implementation of Clinical Analytics in a Digital Hospital. Appl Clin Inform. 2022 Mar;13(2):339-354. doi: 10.1055/s-0042-1743243. Epub 2022 Apr 6. PMID: 35388447; PMCID: PMC8986462.	Seems to focus on digital clinical dashboards from electronic health records, not describing characteristics.
McLennan S, Kahrass H, Wieschowski S, Strech D, Langhof H. The spectrum of ethical issues in a Learning Health Care System: a systematic qualitative review. Int J Qual Health Care. 2018 Apr 1;30(3):161-168. doi: 10.1093/intqhc/mzy005. PMID: 29394354.	Seems to describe ethical issues/risks instead of describing characteristics of an LHS.
Melder A, Robinson T, McLoughlin I, Iedema R, Teede H. An overview of healthcare improvement: unpacking the complexity for clinicians and managers in a learning health system. Intern Med J. 2020 Oct;50(10):1174-1184. doi: 10.1111/imj.14876. Epub 2020 Oct 2. PMID: 32357287.	LHS does not seem to be the main topic, does not describe characteristics of an LHS.
Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. J Clin Oncol. 2012 Dec 1;30(34):4243-8. doi: 10.1200/JCO.2012.42.8011. Epub 2012 Oct 15. PMID: 23071233.	Seems to mainly describe informatics-enabled CER, how CER could interact with a learning health system, and mainly in the field of oncology.
Moloney RM, Tambor ES, Tunis SR. Patient and clinician support for the learning healthcare system: recommendations for enhancing value. J Comp Eff Res. 2016 Mar;5(2):123-8. doi: 10.2217/cer.15.67. Epub 2016 Mar 1. PMID: 26930026; PMCID: PMC5549639.	Seems to describe the paradigm shift towards LHS and its opportunities/challenges.
Price DW, Davis DA, Filerman GL. "Systems-Integrated CME": The Implementation and Outcomes Imperative for Continuing Medical Education in the Learning Health Care Enterprise. NAM Perspect. 2021 Oct 4;2021:10.31478/202110a. doi: 10.31478/202110a. PMID: 34901778; PMCID: PMC8654469.	Seems to mainly concern continuous learning, briefly discusses LHS but does not seem to describe characteristics (rather implementation).

Ramsey LB, Mizuno T, Vinks AA, Margolis PA. Learning Health Systems as Facilitators of Precision Medicine. Clin Pharmacol Ther. 2017 Mar;101(3):359-367. doi: 10.1002/cpt.594. PMID: 27984650; PMCID: PMC5309135.	Seems to concern a network specifically for inflammatory bowel disease.
Richesson RL. Learning health systems, embedded research, and data standards-recommendations for healthcare system leaders. JAMIA Open. 2020 Oct 12;3(4):488-491. doi: 10.1093/jamiaopen/ooaa046. PMID: 33619464; PMCID: PMC7665565.	Formulates recommendations for LHS, does not seem to be describing LHS characteristics.
Rosenthal GE. The role of pragmatic clinical trials in the evolution of learning health systems. Trans Am Clin Climatol Assoc. 2014;125:204-16; discussion 217-8. PMID: 25125735; PMCID: PMC4112713.	Seems to briefly re-iterate IoM workshop findings and thereafter discusses pragmatic trials without the main focus on LHS.
Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, Taylor D, Tibble H, Mossialos E. Health information technology and digital innovation for national learning health and care systems. Lancet Digit Health. 2021 Jun;3(6):e383-e396. doi: 10.1016/S2589-7500(21)00005-4. Epub 2021 May 6. PMID: 33967002.	Does not mainly seem to discuss LHS, rather digital systems and health information technology is discussed.
Slutsky JR. Moving closer to a rapid-learning health care system. Health Aff (Millwood). 2007 Mar-Apr;26(2):w122-4. doi: 10.1377/hlthaff.26.2.w122. Epub 2007 Jan 26. PMID: 17259193.	Seems to be perspectives on steps to take towards an LHS.

ADDITIONAL FILE 3

Table S1 – Evidence table with extracted data for Evidence Ecosystem characteristics

Author (year)	Identified areas/ components in ecosystem, model and/or network	Actors in the ecosystem, model and/or network	Identified links / cooperation in the ecosystem, model, and/or network	Identified products / data in ecosystem, model, and/ or network	Identified concepts and/ or forms of evidence presentation, packaging of information, communications, and/or data exchange	Data quality / development values and/or procedures	Methods for updating / maintenance / currency	Governance	Platforms
Akl 2020 [1]	Collections Clusters Links	User Evidence synthesizer	Links between intervention represent the possible comparisons within a cluster. Links are populated with all relevant studies and a standardized data file.	PICO question (Network) meta-analysis Evidence map	Metadata to allow sub-grouping and filtering according to user needs Standardized data file containing: study characteristics, risk of bias assessment, summary results, individual participant data. Search engine presenting relevant evidence for the PICO	–	A process for matching emerging evidence to the relevant collection or cluster (i.e. a living universe) Artificial intelligence and automation can aid in identifying, coding, and abstraction of data.	–	Web-based platforms
Bogaert 2018 [2]	Generating health data and indicators (by networks where support was provided to develop comparable, standardized and accessible data and indicators for health status and health determinants, health services and health	Researchers Policy makers Citizens Professionals (At national, European union, and international level)	Distributed structure with a central hub which coordinates the operation of distributed facilities or networks. Organizational network: Bringing together networks (national networks and domain specific networks) in health infor	Health information	–	–	–	Three levels are identified: Decision making level (includes an Assembly of members, a Scientific Advisory Board, an Ethics and Privacy Board) / Executive level (includes a Central Executive Management Office, a Consultation Platform) / Operative level (includes a Network Committee) Elaboration: Assembly of members	Virtual and interoperable repository platforms

	systems) Managing health information (using existing data repositories and hosting data sets in a sustainable and accessible manner) Exchanging health information (sharing best practices in health information and mutual learning to work in the same methodological way) Translating health information (optimizing output into knowledge transfer from data to evidence to be used by policymakers, professionals, and citizens.		mation within and between European Union member states (and associated countries). Key players in health information in a national network interact with each other at national level (stimulated by national network) Domain specific networks respond to current priorities and carry out analyses (for which there may not be capacity within national networks) and can cooperate with national network for data-collection and data-analysis at national level					(full decision making power): members and observers Scientific advisoy board (offer advice): scientific experts Ethics and privacy board (offer advice): scientific experts Central executive management office (management, operational and budgetary day-to-day decisions): general director and core team Consultation platform (align with health information landscape): international organisations and commission services Network committee (oversee national scientific activities): representatives of national health information authorities and international research networks	
Boutron 2020 [3]	Hospitals Research funders Regulatory agencies Academia, institutions Professional bodies Politicians Biomedical journals Citizens Local health	Stakeholders Trialists Funders End users Methodologists Clinicians Statisticians	Contact with trialists (e.g. to inform about outcomes) Trialists can post protocols in registries Trialists can share individual patient data (in a secure process at no costs) Trialists are contacted for	A living mapping of all trials Systematic review Network meta-analysis Individual patient data	Interactive data visualizations Descriptive data Forest plots Risk of bias table GRADE summary of findings tables Evidence profiles	Reporting guideline adherence (e.g. CONSORT) Core outcome sets (e.g. by COMET initiative) Risk of bias assessments Intended and realized data sharing	Living systematic reviews (based on living protocol)	Independent steering committee oversees the project	Website

	authorities		missing data						
Carta-bellotta 2019 [4]	Generation of evidence Synthesis of evidence Translation of evidence (integration of evidence into clinical practice)	Industry Patient associations Payers Charities Health professionals Medical specialty societies Researchers Healthcare managers Ethical committees	Evidence synthesis informs the generation of new evidence Evidence translation informs both evidence generation and evidence synthesis	Systematic reviews Clinical practice guidelines	–	Reproducibility of research and clinical practice guidelines (e.g. EQUATOR for reporting guidelines, AGREE) Transparency (Clinical trial registration in trial registries, systematic review registration in PROSPERO) Data sharing Promoting best practices in the production, updating, and dissemination of systematic reviews and guidelines (e.g. by Evidence Based Research Network, Cochrane handbooks, GIN) Identify high priority areas (e.g. James Lind Alliance Priority Setting Partnership)	–	–	–

						Reduce research waste (e.g. REWARD alliance) GRADE approach (evidence hierarchy and clinical practice guideline recommendation formulation) Conflict of interest disclosure and management (GIN)s Use of knowledge translation frameworks (pipeline framework, knowledge-to-action framework, consolidated framework for implementation research)			
Créquit 2020 [5]	–	Editors Funders Patients and public Regulatory authorities Reviewers Non-govern mental organizations	Link/match between all available publications and reports of one trial	(living) Systematic reviews (living) (Network) meta-analyses Protocol Clinical study report	–	Use of new sources for data (Clinical trial registries for additional results and methodology, journals and journal appendices	Living systematic review Living network meta-analysis	–	–

		Physicians Researchers Decision-makers		Case report forms Statistical Analysis plans Individual patient data Routinely collected data Analysis-ready dataset Summary results		reporting protocols, regulatory agencies, pharmaceutical companies) Automated technology (assisting evidence synthesis, AI, machine learning, text mining, data linkage) Crowd sourcing (citizen science)			
Dumitrescu 2006 [6]	–	Producers of information Users of information Experts Decision makers Policy makers Partner organizations	Experts might be commissioned by the network's members to synthesize the best available evidence or to summarize reports. Decision-makers are provided the option to search in a selection of validated sources (web-based)	Evidence syntheses Evidence reports (including a concise summary, an policy options based on local context)	Evidence reports are syntheses of best available evidence written in an easy-to-read style. Structured publication and verification of research output (semantic technologies) Users are mainly interested in information and knowledge that are relevant, specific and fit for their purposes, concise and easy to understand.	–	–	–	–

Elliott 2014 [7]	Health practice Learning health-care systems Primary research and Health big data Linked data repositories Living systematic review Living evidence services Living guidance Decision support systems	–	The container of a systematic review can be opened when data inside are made available. Availability of this data can be used for integration with guideline development and clinical decision support systems.	Protocols Trial registrations Clinical study reports Individual patient data (Living) systematic review Living guidelines Standards Policies Decision-support systems	–	Rich meta-data (for research outputs)	Living systematic review	–	–
Funk 2022 [8]	Health information Information translation tools (visualization and modelling, packaging and synthesis, communication and dissemination, linkage and exchange) Policy cycle (agenda setting, policy formulation, policy adoption, policy mplementation, policy monitoring and evaluation)	Knowledge producers Researchers Knowledge brokers Policy makers	Links between policymakers and stakeholders (those with access to or producers of health information)	Evidence brief for policy Local health messages Memoranda Annual reports Public health reports Research reports Letter to or from governments	Visualization and modelling tools (Graphs / charts / maps; Modelling and simulation) Information packaging and synthesis tools (Evidence brief for policy; Local health messages / memoranda; Annual reports / public health reports / research reports / letter to or from governments) Communication and dissemination tools (Oral presentation; Newsletters; Email messages; Health information websites / interactive data platforms / surveillance	–	–	–	–

					platforms) Information linkage and exchange tools (Dedicated group of stakeholders; Deliberative dialogue; Knowledge networks; Knowledge brokers) Translation mechanisms: push, pull, exchange, integrated				
Gough 2019 [9]	Use of research for undertaking analysis of situations and informing decisions Engagement of users and producers of research, providing access, and supporting interpretations of research and its uptake Production of both primary research and reviews of such research Other information used to interpret research findings Broader socio-political context withing the ecosystem exists	Evidence producer Evidence user Non-academic users Policymaker Authors of systematic reviews Stakeholders Topic specialists Knowledge brokers	Producers and users of research are very varied and engage with each other in many ways. Authors of systematic reviews may, for example, be skilled methodologists and yet bring in other stakeholders such as topic specialists, likely users of a review, and broader societal perspectives to advise on the questions and methods of a review. Systematic review authors may bring in topic specialists, users, Mutual exchange	Systematic reviews (Network) meta-analysis Qualitative comparative analysis Individual participant data Multi-component reviews Maps Gap maps Maps of maps Reviews of reviews Scoping reviews Rapid reviews	–	Clarity about the standards of synthesis evidence: the method of undertaking, the studies included in the review, the totality of evidence produced Reporting standards GRADE Evidence-to-decision framework A functional ecosystem would ensure that research is curated to make data easy to find, access, and reuse. Review data need to be shared (to prevent dupli	Living systematic reviews Living guidelines	–	A repository of all relevant trials classified according to an ontology (e.g. developed by Cochrane on the PICO model) can be used by organizations to describe the included trials and share data among each other. The repository may include data-extraction,

	of ideas between policy-makers, practice, research, personal decisions, and researchers: decision-makers seeking evidence / researchers offering evidence / both seeking and offering Interpreting research evidence in order to make recommendations or guidance for policy, practice, or individual decision-making. Integrate research evidence, complexity perspectives, and norms and values of decision-makers and stakeholders Supply-driven reviews (clarity and consensus about definitions and concepts provide a sound starting-point) Demand-driven reviews invest more resources early on for understanding problems faced by users (engage directly with those users).	cation).	risk of bias assessments, selection of outcomes, and interpretation of findings.	

Jordan 2022 [10]	Global health (sustainable impact, engagement, knowledge need) Evidence generation (research, expertise, discourse) Evidence synthesis (systematic review, evidence summary, guidelines) Evidence transfer (active dissemination, systems integration, education) Evidence implementation (context analysis, facilitation of change, evaluation of process and outcome) Overarching principles (culture, capacity, communication, collaboration)	Health professionals Patients Knowledge creators Knowledge curators End user Stakeholder	Critical relationship between research and practice: research and synthesis sectors need to be engaged with those working in policy or practice and patients Uptake of evidence at the point of care by active disseminations, systems integration and education Bidirectional arrows between the components indicate fluidity of the form and function of eacht of the component parts (global health, evidence generation, evidence synthesis, evidence transfer, evidence implementation) Cooperation between key stakeholders in the evidence ecosystem is critical for multifaceted approaches and pragmatic solutions in the generation of trustworthy evidence across sectors Communication,	Systematic reviews (and 11 types of evidence syntheses: effectiveness, qualitative, text and opinion, prevalence and incidence, economic, etiology and risk, mixed methods, umbrella reviews, diagnostic test accuracy, scoping reviews, measurement properties review) Evidence summaries Clinical guidelines (and related derivative content)	–	Guidance about 11 types of evidence syntheses: effectiveness, qualitative, text and opinion, prevalence and incidence, economic, etiology and risk, mixed methods, umbrella reviews, diagnostic test accuracy, scoping reviews, measurement properties review There are opportunities for new technology and artificial intelligence to improve efficiency and reliability of outputs.	–	–	–

			collaboration, and integration of evidence: coactive and participatory communicative process between creators (and curators) and end-users of evidence.						
Love-stone 2020 [11]	Data harmo-nization via a common data model A platform to find access and (re) use health data Governance to comply to Euro-pean (privacy) laws and access data	Researchers Data-custo-dians Data owners Academics Subject matter experts Patient organi-zations Pharmaceu-ti-cal companies Statisticians	Research-ers browse meta-data in the catalogue to data suitability and feasibility of a particular study. After approval of the data owner(s), the study may be conducted in a secure analysis environment	Electronic health record data Research cohort data High-level meta-data	Data-visualization (on the EMIF platform) A database taxonomy based on meta-data and a data dictionary was created Common data model (for harmo-nization to stan-dardize the format and content of observational data [population and cohort data])	An in-house minimal data set for cohort data was developed based on the common data model	–	An ethical code of practice (to ensure data protection, patient confiden-tially, and compliance with Euro-pean privacy laws) was developed for ethical governance. Tools in the EMIF-plat-form were able to run locally at the data-custo-dian's site. This provided local provenance and governance. Only if required and with approval of data-cus-todians, data could be exported to external software (e.g. SAS, R, Excel) The EMIF-PLAT devel-oped a Code of Practice (known as the ECoP) to ensure privacy protection of data subjects and to protect the interests of all data-sharing parties.	EMIF-plat-form to find, assess and (re) use health data from diverse sources across Europe. It contains a set of tools designed to enable data-dis-covery and charac-ter-ization. Additional modules were de-veloped for documen-tation, task manage-ment, process manage-ment, and analysis of the

requested data in a secure environment. It is based on a federation of databases, rather than a single centralized database. Tools in the platform: catalogue, federated analyses of electronic health records, patient selection tool, variable selection tool, task management, private remote research environment, codemapper The tranSMART platform was used as in integration platform where data was harmonized to a newly established data set

									and stored on the platform.
Mathews 2019 [12]	Evaluate and improve policy and clinical practice Produce evidence Synthesize evidence Knowledge translation	Evidence producers Evidence synthesizers Evidence processors Disseminators Evidence implementa-tors Decision makers Health care providers Patients The public Program managers Policy makers Those involved in delivering and using health services Stakeholders	Up to date (global and local) evidence needs to flow efficiently between producers, synthesizers, processors, implementors. The flow through the ecosystem is not unidirectional. Local circumstances (e.g. feasibility, burden of disease) may influence the entry point into the flow. Close working relations between policy-makers and researchers (e.g. commission research Policy-maker pull in relation to evidence: demand-driven research commissioned by policy-makers or other stakeholders)	Primary health systems research Evidence syntheses Decision support products (guidelines, guidance, policy briefs, evidence summaries, gaps in primary research) Quantitative data from RCTs for effectiveness Economic evidence for resource-use (what resources are needed to achieve benefits and hoe they should be prioritized) Qualitative evidence for stakeholder views, acceptability, and feasibility.	Packaging review findings and guidelines into user-friendly products (e.g. SUPPORT summaries), to facilitate the flow of evidence into policy processes.	Use of platforms that incorporate trustworthy standards for decision support systems, such as GRADEpro and MAGI-Capp	–	Governance of evidence through an independent quality assurance framework (including a process for declaring potential conflicts of interest)	A range of platforms for synthesis, such as Covidence and Cochrane Review Manager. Platforms to process evidence, such as GRADEpro and MAGICapp that facilitate data sharing across the evidence ecosystem.
Pilla 2022 [13]	Functions: Community building (build a shared	Entities (groups and individuals)	The relationships and interactions within the	–	–	–	–	Governance referred to operational structures, roles, responsibilities,	–

vision and cohesive, supportive relationships among diverse stakeholders by increasing transparency and trust) / Knowledge management (acquire, develop, exchange, and disseminate knowledge) / Amplification and advocacy (places issues on global agenda, amplifies member voices, enhances status and legitimacy of individuals and the network) / Convening (developing meaningful connections between diverse groups, foster consensus, and capacity building through shared experiences) / Resource mobilization (resource mobilization capacity for long, medium, and short-term initiatives)
Form (JBI Collaboration form: membership, relation

JBIC form the substance of the network.
The relationships between members of the JBIC (with each other and with JBI) are multifaceted, ranging from structured, formal relationships with goals and objectives (e.g. collaboration agreements, mentorship, clinical partnerships), to social relationships cultivated over years of shared experiences, with the JBIC referred to as a “global family”.
The JBIC is a centralized network, whereby the global architecture is built around a governing entity (JBI) with “nested” groups organised by region, managed structurally with regional chairs to foster regional collaborative efforts and mentorship and support of novice groups by more

and decision-making. An operational framework was co-created to support a shared vision and mission. Including the recognition of important scientific, scholarly, and collaborative activities by entities in the collaboration.

	ships, benefits, governance, organizational arrangements, stewardship, resources)		experienced ones.						
Ravaud 2020 [14]	Continuous refreshing of evidence syntheses Encouraging communities to work together and cocreation of evidence syntheses Tailoring end products to the stakeholder needs Optimizing trade-off between speed and thoroughness Research on evidence synthesis methods Priority-setting and global distribution of work Living evidence communities	Researchers Trialists Stakeholders (clinician, researchers, policymakers) Health professionals Guideline developers Patient (representatives) Methodologist Statisticians Health system managers Funders Journals Universities Regulators Governments Health technology assessment agencies Health policy authorities	Link between primary evidence and evidence synthesis (e.g. sample size calculation for a trial informed by a meta-analysis, recording of all important outcomes) Cocreation to help producing tailored outputs Distributing workload for evidence syntheses among teams at the international level Bidirectional interactions with stakeholders Feedback loops between living evidence synthesis production and primary research communities: Live monitoring of trials conducting quality (outcomes, risk of bias, before conducting a trial), live monitoring of trials	Rapid reviews Evidence syntheses (Living) systematic review (Network) meta-analysis Observational data Trial data Randomized controlled trials Real-world data Living mapping of research	Standards for reporting data to bypass data-extraction and include data from primary producers straight to syntheses. Products should have varying formats and sizes, different focus of content, different language complexity, and could be translated to different languages.	Core outcome sets Cochrane risk of bias tool Post results in a trial registry (transparent and complete reporting) Give access to trial protocol Share trial data Living disclosure of the quality, transparency, and accessibility of data (to motivate actors to develop quality improvement programs about reporting transparency)	Living systematic review	The ecosystem must be organized at a global level with a geographically wide-spread reach, rather than a geographically localized distribution of work. Safeguards should be in place to avoid undue influence of lobby bodies.	A marketplace for evidence syntheses, where stakeholders will find evidence they need on a specific topic. A practical repository solution to share trial data.

transparency (transparent reporting of trial report, protocol and data), living disclosing the quality and transparency results (of quality / transparency / accessibility of data), living mapping of research (to find gaps and support researchers planning a project, funders, and institutional boards)
Guideline developers find the data from living network meta-analyses on a marketplace and produce locally relevant guidelines
Implementing living guidelines and living network meta-analyses will help outline what is currently unknown. This helps the research community to streamline future clinical trials for where there is deficient evidence
Create relation

			ships between evidence synthesis community and guideline developers, regulators, and governments to support, promote and sustain living evidence synthesis.						
Shepherd 2014 [15]	Evaluation funders Evidence generation Evidence pumps Storage (publication, open access, quality control) Evidence synthesis, guideline production and pumps Evidence user Push forces are prominent in the evidence generation, where pull forces are more active in the evidence users	Funders Policy-makers Commissioners (General) practitioners Clinicians Practitioner-academics Knowledge translators (evidence synthesizers who consolidate and 'pump' evidence in accessible and usable formats) Specialist communities in medicine / professional bodies Government Research councils Research charities Non-governmental organizations	A system cannot function without conduits connecting each stage. Pipelines are used for transport from storage facilities to distribution points (interfaces with end users). Product blending (e.g. evidence from process evaluations might be very useful blended with effectiveness data). Some evidence might be more difficult to 'pump' in the ecosystem than others. Evidence needs to be synthesized and translated into policy and practice guidelines. Academics and practitioners in health care are	Systematic review Health technology assessment appraisals (costs, effectiveness, boarder impact of health care) Multi-site/ single-site RCTs Quasi-experiments Process evaluations Qualitative information Participants and program staff anecdotes and observations	Evidence is often categorized according to the evaluation method (randomized, quasi-experimental, qualitative, etc.) but can also be classified according to the practitioners' and commissioners' need for usable forms. Evidence refinement is likely to be needed, as evidence production and synthesis will be wasted when the usability is low. Knowledge translators should synthesize evidence and promote it in short summaries.	Waste, unusable evidence, and contaminated evidence (e.g. statistical jargon, wrong methods, missing promotional opportunities for findings) need to be recognized and minimized To be implemented, evidence needs to be timely, usable, clear, relevant, credible, and generated using uncontested methods. A healthy ecosystem generates evidence using appropriate	–	An evidence ecosystem contains many parts. System sensors are needed so that faults are identified and put right promptly.	Storage points (interfaces with end users)

Universities
Media
Business and commerce
Technology strategy board
Campaigners
Guideline developer

often the same people. This may overcome the problem that guidance and policy do not reflect practice.
For clinicians: journals in hard copy aimed at their specialties, group discussion, and commissioning teams to summarize evidence.
Guideline developers work closely with funding bodies to support evidence generation
Integration of a guideline developing unit in a specialist community provides an incentive to engage with and apply evidence.
Institutions and roles which provide more than one ecosystem functions help to connect the entire system.
Evidence and cost-benefit need to be applied in the context of the setting. Such evidence is generated by action research methods
Research assistants in commissioning teams should be targeted for evidence skills training.

			and should be commissioned or sought by commissioners. Evidence needs to flow through the ecosystem from generation to end user. It will not flow on its own. “Product push” and “demand pull” are needed at every stage.						
Stewart 2019 [16]	Structural foundations (political and constitutional drivers) Organizations (key organizations within the ecosystem Investments (resources) Capacities (on individual and organizational levels_ Innovations (adaptation and responses to change) Five dimensions of complexity throughout the ecosystem: complexity in individuals and organizations within the system (role-players), complexity in the activity	Government departments National treasury Advisory level Funders / funding agencies Producers Users Statutory science councils Public universities Higher education institutions Researchers Decision-makers Policy-makers Knowledge brokers	Researchers are responding to national priorities and strategies Government and research community work together (formed a group of evaluators) Academic centers and departments focus on: production of research to inform the achievements of the National Development Plan, access to research to inform development decision-making, understanding of the policy and planning process, the collection of evidence for	Performance expenditure reviews Policy relevant evidence mapping Responsive evidence synthesis Evidence (gap) map Systematic reviews Impact analyses Monitoring data Administrative data Citizen monitoring data Citizen science Process evaluations Large longitudinal datasets	Visually appealing outputs	–	–	The government had a monitoring and evaluation system in place (including citizen monitoring through citizen report cards to monitor service quality) The Department of Public Service Administration have been driving strategic and operating standards for public service administration.	A secure data facility where administrative data is prepared and curated for research and policy analysis. South Africa’s Sustainable Development Goals Knowledge Hub as a central portal of repositories from different universities organized round different sustainable development goal

	spectrum from evidence production to use (activities), complexity of sectors involved in evidence-informed decision-making (sectors), complexity in evidence types (types), complexity in the decision-space for which evidence is sought (range of questions).		decision-making. Organizations driving both the demand and supply of evidence. Co-production and networking bring internal and external initiatives closer.						themes. Specialist systematic review software from international organizations (e.g. Cochrane, EPPI-Centre) for efficient management of large amounts of research. The Africa's Centre for Evidence and Department for Planning, Monitoring and Evaluation evidence mapping software applications for visualizations
Vandvik 2020 [17]	Produce evidence Synthesize evidence Create trustworthy guidance Disseminate to policymakers, clinicians, and patients Implement evidence and	Clinicians Patients Methods experts Systematic review teams	Links with actors producing evidence upstream and those implementing evidence and evaluating impact in practice downstream. Create a link between living evidence synthe	Primary research Real-world data Systematic reviews Digitally structured guidelines HTA and (interactive) decision aids	Digital, multi-layered and user-friendly HTA-reports Guidelines and decision-aids on all devices and in web-portals Ready for re-use, adaptation and plug-ins. Interactive	–	Rapid recommendations to overcome organizational barriers in updating guidelines Living evidence	Need for an overarching infrastructure to, govern, actors working across the steps Strict management of conflicts of interest and patient partnership	MAGICapp interoperable platforms for systematic reviews, guidelines, and clinical decision support systems in electronic

guidance Evaluate and improve practice Core requirements to make an evidence eco-system function: Common methods, Coordinated processes, Tools and platforms, Digitally structured data, Culture for sharing and innovation, Global standards, Orchestration / governance / support	sis community, researchers, and decision makers for specific clinical questions and living network-meta-analysis Clinicians, patients, methods experts co-create guidelines, coordinated with systematic review teams. Evidence synthesis are linked to guidelines through common standards, methods, processes and platforms. Data can flow between interoperable platforms for systematic reviews, guidelines, and clinical decision support systems in electronic health record. Downstream decision support could be linked to active implementation of evidence into practice and health care systems	Evidence summaries from reviews Clinical decision support systems Electronic health records Care plans Population-based data in registries Quality indicators Network meta-analysis	infographics with direct links to multi-layered digital content Interactive decision aids			health record.

Table S2 - Evidence table with extracted data for Learning Healthcare System **characteristics**

Author (year)	Identified areas/ components in the learning health system	Actors in the learning health system	Identified links / cooperation in the learning health system	Identified products in learning health system	Identified data sources in learning health system	Identified concepts and/or forms of evidence presentation, packaging of information, communications, and/or data exchange	Data quality / development values and/or procedures	Methods / mechanisms for learning or evaluation	Governance	Platforms
Apfeld 2020 [18]	From a reference [6]: Seamless knowledge sharing Transparent and continuous evaluation of interventions Generation of new evidence to reduce clinical and system uncertainty From a reference [7]: An LHS combines research, data science, and quality improvement to yield knowledge as a by-product of the patient-clinician interaction (aimed at improving patient health and system outcomes)	Clinician-scientists Primary care givers Patients Family members Health system leaders	Clinician-patient interaction From a reference [9, 10]: Patients, clinicians, family members, health system leaders are working together to coproduce healthcare, sharing expertise and experiences	–	Electronic health records Use-case cohorts	–	–	–	–	–
Atkins 2017 [19]	Data and access to evidence-based knowledge are key foundations of a learning health care system IoM model: Real-time access to knowledge Digital capture of	Patients Health care team Clinicians Researchers Clinical program leaders	Definitive findings from well-designed clinical trials shape practice through their incorporation into guidelines, quality measures, or other system-wide implementation efforts. The evidence base	Peer support programs	Clinical and administrative data in corporate data warehouse Annual survey data on patient experience Clinical	–	Independent researchers publish on the quality of care	Strong academic affiliations Clinicians with dual appointments Strong emphasis on system redesign and	–	Patient portal

care experience
Engaged, empowered patients
Incentives aligned for value
Full transparency
Leadership-instilled culture of learning
Supportive system competencies

is insufficient to address current needs, researchers and clinical program leaders collaborate to establish the evidence, through common research priorities and data collection strategies, and use research to inform progressive iterations of new clinical programs.
Policy changes and clinical innovations driven by the health care system serve to generate evidence via natural experiments that can be used to test specific strategies.
Electronic health record gives access to guidelines, clinical reminders and alerts
Home telehealth to monitor patients
Patient email communication with health care team (through patient portal)
Patient access to medical records, clinician notes, and educational information (through patient portal)
Peer support to promote engagement among patients
Clinicians have access to clinical dashboards

learn training
From reference [14]: continuous testing of interventions in real-world settings, rather than relying on efficacy trials in controlled settings

			dashboards to compare their quality relative to peers							
Britto 2018 [20]	A repository is created and shared with information, knowledge, resources, and know-how. From a reference[16]: Aligning participants around a common goal Standards, processes, policies, and infrastructure enables collaboration Governance and policies Network management Quality improvement Research facilitation Engagement Information technology	Patients Clinicians Researchers Families Medical center People, organizations, databases, registries (actors in an organizational architecture) Network leader High-performing sites Staff	Patients and clinicians work together Stakeholders (patients, families, clinicians, researchers) participate in all aspects of governance. Research prioritization involving stakeholders Training resources for multiple stakeholders	Observational research	Registries Electronic health record	Quality improvement tools: logic model to describe and document theories of change	Standardized processes, protocols and policies (including: communication policies, data-sharing, privacy protection, regulatory compliance) Standardized technology used to reduce data-entry burden, facilitate care, which results in data useful for research and learning. Participants commit they will not use safety data for competitive purposes or display it publicly Building a safety culture Standard operating procedures Standard framework	Collaborative quality improvement using a modified Breakthrough Series approach Integrating subject and improvement knowledge Ongoing learning Alternating learning sessions and action periods (structured) Statistical process control methods Common repository to share resources created by participants. Shared knowledge Shared tools and resources (e.g. registry data reports) Shared standards (e.g. common data model, data transfer standards, measurement standards, control chart	Network governance processes: Membership policies (guidelines, rights, obligations) Collaboration and attribution policies (authorship, copyright, intellectual property) Data sharing policies (access, ownership, privacy, security) Research and regulatory policies (informed consent, master reliance agreement) Privacy oversight (patient generated data)	–

for quality improvement

standards.
Shared situational awareness (i.e. problems and opportunities in the organization's environment)
High degree of transparency and ongoing sharing of measures and aggregate and site-level performance (outcome and process outcome).
Observe own results and compare them to others, learn from others' experiences
Sharing ideas, knowledge, and know-how: 'all teach, all learn' culture
Networks evaluate their policies in regular cycles
Evaluate management structure regularly
Research prioritization involving existing evidence,

								and network outcome/process data. Evaluate the impact of research resources Monthly calls and meeting to evaluate the network functioning, highlighting areas where learning can be harvested		
De Bruin 2022 [21]	Financial incentives Data-infrastructure Policy-infrastructure Technology Data Learning Evidence and measurement Culture (change) Stakeholders	Patients Family members Patient advocates Care providers Clinicians Payers Policy makers Healthcare administrators Researchers Technology experts Health system leaders Thought leaders on continuous improvement Health service managers Planners	Interaction between clinical practice and research domain (accelerating research and implementation of knowledge in practice) Interaction between stakeholders (for joint goal-setting, revealing underlying values, evaluating processes, share best-practices)	Data-dashboards Prognostic models Clinical decision support tools	Data flowing from routine care Health-related data (Not specified) PROMS Experience of care (patient and professional) socials determinants of health patient generated data (electronic health records (EHRs), claims databases, pharmaceutical clearinghouses, and clinical trial	data infrastructure was interpreted as the linkage and/or storage of different data-sources (support system)	–	Learning as (intermittent) information exchange between the clinical domain and the research domain Learning as a (technology-aided) continuous circular process of converting (routine care) data to knowledge; knowledge to performance; performance to data [central is the information stream] Learning as a recurrent interaction	–	Support system

					databases) geospatial data			between stakeholders to reveal/ discuss underlying values, to evaluate processes, and to identify opportunities for change and share best-practices. Human interaction and the exchange of experiences and ideas are central.		
Budrionis 2016 [22]	–	Patients Doctors / clinicians Service providers Health system stakeholders Health system leaders	–	Prediction models Large pragmatic trials	Data collection in clinical practice Electronic health records Patient-reported measures / outcomes	Interoperability: comprehensive data models and semantic interoperability mechanisms	–	Quality improvement through peers changing each other's behaviour based on historical performance. Defining the problem, developing algorithms, implementing tools, evaluating performance of the tools and their effect on patient care and outcomes. Evaluate the effects of implementation to redefine or de-implement ineffective or	–	Networks of nodes, locally processing sensitive patient data and sharing products of computation by coupling open-source components. Or, alternatively:

								harmfull interventions.		centralized architecture
Casey 2021 [23]	–	Patients Clinicians Health system leaders (external) Researchers Funders Industry representatives	Stakeholder engagement	Pragmatic trials Comparative effectiveness research (levels: patient, unit, hospital) Trials using information technology to identify patients likely to experience an outcome or benefit from an intervention Prognostic and predictive models Trial evaluating models of healthcare delivery	–	–	–	–	–	–
Deans 2018 [24]	From a reference [5, 8]: Coproduction of healthcare: patients, clinicians family members, healthcare leaders working together as partners.	Patients Clinicians Family members Health system leaders Researchers	Clinician-patient interaction Patients, clinicians, family members, healthcare leaders working together as partners Active collaboration from all partners in the system Experienced and agile teams of software engineers, database engineers, extract, transform, and load (ETL) experts, computer sci- entists,	–	Research studies/ biomedical research Electronically captured and combined data from patient-clinician interaction	–	–	Research influences practice, practice influences research. Point of care -> research during normal clinical care -> integration in knowledge networks (including biomedical research)-> application of best available	–	Software engineers, database engineers, extract transform load experts, computer scientists,

			informaticists, and analysts create the digital infrastructure necessary to stand up a LHS framework. However, equal commit- ment on the part of clinicians, quality improvement specialists, healthcare system leadership, researchers, and patients is neces- sary for the LHS to fully realize its potential.					evidence (taking in to account patient needs, local system, community context)		informaticist, analysts create a digital infrastructure.
Easterling 2021 [25]	Organizational learning, innovation, and continuous quality improvement Quality improvement processes are standard practice (continuous improvement processes are used, Rapid learning processes are used, Plan-Do-Study-Act cycles are used) Learning is done according to particular principles, processes, practices and/or models (collaborative team-based learning, Systems science, Learning cycle [Friedman], collective sensemaking, positive deviance, triple-loop learning, emergent learning) Learning is driven by specific goals (equity,	clinicians, patients, and other stakeholders Family members Researchers community-members, community-based organization, payors	clinicians, patients, researchers, community-members, community-based organization, payors and other stakeholders are actively engaged in the learning process and/or clinical decision-making and/or improving the organization and/or research.	–	–	–	–	Organizational culture emphasized and supports learning (transparency is values by the organization, collaboration is valued, culture facilitates trust-building) Learning is championed by organizational leaders (clinicians are encouraged to conduct research, all employees are expected to be active learners)	–	–

improving the quality of care, efficiency, patient safety)
Learning takes place throughout the organization
Translating knowledge and evidence into improved practices:
Research is translated into practice (research within the organization is translated, research findings from the literature are translated)
The organization adopts or implements evidence-based treatments
There is a reciprocal relationship between research and practice
The organization is systematic in its implementation processes (interventions should be adapted or tailored to the specific context, allowes dot learning and refinement in implmentation, systematically de-implements practices that no longer serve the organization, follow principles of implementation science, implements with fidelity)
Building new knowledge and evidence:
The organization buidls knowledge or

evidence
The organization conducts research
The research conducted by the organization is practical or needs to balance practical with rigorous
Findings from the research are shared or disseminated (internally, externally)
Research conducted by the organization answers questions that are directly relevant to the organization (are questions posed by clinicians relevant to clinical practice, answers questions by organizational leaders relevant to organizational goals)
Data are translated to information
Internal knowledge and external knowledge are integrated
The research conducted by the organization is rigorous
Analyzing clinical data:
Patient data are captured and organized into a system, which is then used for analysis
Clinical and/or informatics data are used in diagnosing and treating individual

patients (clinical decision support systems are in place, personalized treatment e.g. using genomics data, precision medicine)
Aggregated clinical data is shared between institutions (clinical data systems of different institutions are networked)
Clinical data are analyzed to develop research questions and design studies
Engaging clinicians, patients, and other stakeholders:
Patients and family members are actively engaged (in learning process, in clinical decision-making)
Stakeholders beyond researchers are engaged in the learning process (stakeholders from within the organization [beyond researchers] are engaged in the learning process
Community members or community-based organization are engaged (in the learning process, in improving the organization)
Clinicioans are actively engaged in research
Payors are engaged in the learning process

Ether-edge 2007 [26]	–	Policy-makers Research epidemiologists Researchers Physicians Patients Sponsors Research networks Federal government Private sector	–	Knowledge gaps	Electronic health records (computer-searchable) Databases Clinical trial databases	–	National EHR data, registry, connectivity, reporting, and privacy protection standards.	Generate and test hypotheses Learn from large numbers of patient experiences through EHR systems.	–	EHR search software to provide more relevant information for individuals (rather than group averages).
Guise 2018 [27]	National-scale practice-based networks LHS to generate their own data LHS to share the results of their own data Multi-level: Single delivery systems Collaboratives	Authors Journals	From references [6, 7]: Producers of systematic reviews tailor the review process and presentation of results to facilitate brining evidence to practice.	–	External evidence from trials, studies, and reviews	Representing evidence in computable forms: Standardized representation Computable representation Text, tables, figures Evidence can be packed as digital knowledge objects using a standardized format, including the addition of meta-data.	–	Performance to data -> Data to knowledge + Critically appraised external evidence -> Knowledge to performance -> etc.	–	Digital knowledge objects can be curated and managed in digital libraries
IoM 2011 [28]	Culture Participatory Team-based Transparent Improving Design and processes	Private insurers Regulatory agencies Clinician	Basic data exchange infrastructure: Physicians, hospitals, labs, pharmacies generate health	Registries, Health plans Nationwide health information network	Electronic health records Personal health	Meaningful use requirements: Core structured personal	Next generation requirements: Technical progress		–	Health information portals

Patient-anchored and tested
Patients and public
Fully and actively engaged
Decisions
Informed
Facilitated
Shared
Coordinated
Care
Starting with best practice (every time)
Outcomes and costs
Transparent
Constantly assessed
Knowledge
Ongoing
Seamless product of service and research
Digital technology
Engine for continuous improvement
Health information
Reliable
Secure
Reusable source
Data utility
Data stewardship
Used for common good
Trust fabric
Strong
Protected
Actively nurtured
Leadership
Multi-focal
Networked
Dynamic
Common themes and principles in the LHS:
• Build a shared learning environment
• Engage health and

Physician
Hospitals
Labs
Pharmacies
Stakeholders:
Federal
Health and human services [coordination of federal efforts, program management, content development, public health monitoring, implementation of health information tech, post-market drug/device-surveillance, data reuse, improve access/ coordination, collaborative research and rapid translation from study to clinic]
Governmental departments [standards and interoperability, telehealth research, funding of digital infrastructure R&D,

information at each patient encounter
Health information is stored (registries, health plans, nationwide health information network, health information exchanges)
Stored data is used for electronic feedback (patient lists, decision support, alerts)

Health information exchanges

records
Telehealth
Electronic monitoring devices
Biobanks
Health information databases

data (e.g. age, sex, ethnicity)
Core list of active problems
Core structured clinical data (medication, lab results)
Clinical decision support
Care coordination support/ interoperability
Outpatient medicines electronically prescribed
Automated medication safeguard/ econciliation
Visit-specific information to patients
Automated patient reminders
E-record patient access (e.g. copy, patient portal)
Embedded clinical quality measures
Security safeguards
Condition specific data retrieval capacity
Public health reporting for

Ultra-large scale system perspective
Distributed, local data maintenance
Virtual interoperability
Reliable use, systems security protocols
Standards for setting, metadata, vocabulary, data transport, common core datasets, sentinel indicators, access authorization, data quality review protocols
Knowledge generation and use
Core clinical data elements available for quality improvement and research
Channels and protocols for integrating clinical and public health

health care, population and patient
• Leverage existing programs and policies
• Embed services and research in a continuous learning loop
• Anchor in an ultra-large-scale systems approach
• Emphasize decentralization and specifications parsimony
• Keep use barriers low and complexity incremental
• Foster a sociotechnical perspective, focused on the population
• Weave a strong and secure trust fabric among stakeholders
• Provide continuous evaluation and improvement

electronic/ personal health record system design and use for care and research]
IT companies [software to support clinical an business functions, patient portal]
Health care delivery organizations [use digital capacity to improve care, increase aptient involvement, sped research insights]
Academic medical centers [to speed research insights from clinical care, and apply research findings to improve clinical care]
Cooperation capacity resource organizations [implement and use data-sharing and

reportable conditions
Advance directives for ages over 65

Capacity and protocols for query-driven data use in quality, research, and monitoring of sentinel indicators
Novel statistical and database tools for reliable new insights
Patient and population engagement
New norm for patient involvement
Facilitated personal record interface
Clinician-patient electronic partnership
Patient information access and control
Updated best practices delivered ad point of decision
Active patient support for data use in care improvement

	distributed datasets] Stakeholder organizations [advance stakeholder interest in health care tech development and use] Independent sector [funding and facilitation of innovation]		Clinician-public health e-partnership Governance Progressively evolving requirement, specifications, process protocols for exchange, interoperability, research Cross-national harmonization Broad ongoing evaluation capacity

IoM 2011 [29]	Culture Participatory Team-based Transparent Improving Design and processes Patient-anchored and tested Patients and public Fully and actively engaged Decisions Informed Facilitated Shared Coordinated Care Starting with best practice (every time) Outcomes and costs Transparent	Patient Clinicians Patients' families Research Healthcare stakeholders Communities	Patient-clinician interaction Patients engaged as full research partners Science generates evidence that informs and shapes care delivery	Registry functions Decision support tools Medication alerts Best-practices	Clinical data (EHRs)	–	Health care must depend on individuals who can be trusted without conflict to provide science-based advice and act for the well-being of the people they serve. Careful and constant attention to privacy and security issues will be essential to securing such as	Relevant, real-time results to enable more effective care, encourage participatory and science-based decision-making, and foster continuous learning for all healthcare stakeholders. Data and information -> analyzed -> problematic factors identified -> action identified -> actions	–	–

Constantly assessed
Knowledge
Ongoing
Seamless product of service and research
Digital technology
Engine for continuous improvement
Health information
Reliable
Secure
Reusable source
Data utility
Data stewardship
Used for common good
Trust fabric
Strong
Protected
Actively nurtured
Leadership
Multi-focal
Networked
Dynamic
Patient anchored care:
Listening [each patient-clinician interaction starts with uninterrupted attention to the patient's voice on issues, perspecives, goals, preferences]
Participatory [patients engage in their own care]
Reliable [patients should expect proves best practices at the starting-point]
Personalized [science-based tailoring informed by personal biological traits,

public trust and support. Transparency means publishing data with the intent of improving government accountability and enabling citizens to access and use data in ways that are socially beneficial. Public participation means welcoming and accessing widely dispersed national interests and expertise to help the government better accomplish its mission of improving the well-being of the American people. Collaboration means the federal government will work effectively across agen

assigned, performed, monitored -> feedback to clnicians, employees, leadership -> validation -> data and information -> loop continues.

circumstances, and preferences]
Seamless [care by multiple providers in multiple settings should be fully integrated and seamless]
Efficient [Patients, families, and clinicians should expect care appropriate to the need, resources and time required]
Accountable [all relevant aspects of the clinical experience should be captured and routinely assessed against expectations]
Transparent [outcomes of care should be accessible and understandable to patients and their families]
Trustworthy [strong and secure foundation of trust on safety, quality, security, efficiency, accountability, equity]
Learning [the patient is an active contributor to and supporter of the learning process]
Knowledge-driven healthcare delivery system:
Patient-centered care [focus on quality and coordination of care]
Real-time data and

cies; with state and local governments; and with others outside of government, nonprofits, advocacy organizations, business, and academia.
Open government
Ensuring Open Access to Quality Information

	feedback for providers at point-of-care Culture of collaboration, innovation, translation of scientific knowledge into improved health for patients and communities Health technology information systems [integration, standardization, interoperability] Delivery of high-value healthcare in an information-enabled single practice									
IoM 2013 [30]	Science and informatics Real-time access to knowledge Digital capture of the care experience Patient-clinician partnerships Engaged, empowered patients Incentives Incentives aligned for value Full transparency Continuous learning culture Leadership-instilled culture of learning Supportive system competencies	Patients Families Other caregivers Clinicians Communities Consumers	–	–	Clinical data RCTs Observational studies SR Medical product research	–	These strategies, including lean, Six Sigma, and others, introduce methods for coordinating complex work across diverse organizations, identifying existing and potential problems, and addressing those problems systematically. All of these strategies imply that the goal should not be to make	A learning healthcare system constantly refines care operations and processes through ongoing tea training, skills building, system analysis, information development, and creation of feedback loops for continuous learning and system improvement Leadership, incentives, and culture support the continuously	–	–

learning healthcare system for systematically capturing and translating information generated by clinical research and care delivery.

the system work perfectly immediately, but to establish a process of gradual improvement. Improving quality and controlling costs requires moving from this unsustainable and flawed organizational arrangement to a system that gains knowledge from every care delivery experience and is engineered to promote continuous improvement. Transparency of process, outcome, price, and cost information, both within health care and with patients and the public, has untapped potential

							to support continuous learning and improvement in patient experience, outcomes, and cost and the delivery of high-value care.-> reporting systems			
IoM 2013 [31]	–	–	–	Large simple trials	Electronic health Records PROMs	–	The used methods need to be rigorous and the evidence must be valid and generalizable	–	–	–
IoM 2013 [32]	–	Clinicians Patients Decision makers Researchers Scientific community	–	Observational studies (on the effectiveness of therapies in the real-world clinical practice), including cohort / case-control / self-controlled case series Randomized controlled trials Predictive models Observational studies in conjunction with RCTs	Electronic health records Mobile sensors	–	–	Leadership, incentives, and culture support the continuously learning healthcare system for systematically capturing and translating information generated by clinical research and care delivery.	–	–
IoM 2015 [33]	A learning health system aligns	Providers Patients	–	Decision support systems	Electronic health	–	–	Complete learning cycle	–	A platform

	science and informatics, develops strong patient-clinician partnerships, provides incentives for innovation, and creates a culture of continuous improvement to produce the best care at the lowest cost (Steve Leffler)	Researchers Learning networks Specialty societies Patient groups Hospital groups / hospital systems Accountable care organization Foundations Research centers Academic health centers Community centers		Health plans	records Claim information Distributed but networked databases Clinical history Lab samples Genomic data			(Friedman, IOM workshop 2014): Decision to study [mechanisms for communities of interest to form] Assemble relevant data [policies governing access to data] Analyze data [technology for aggregating and analyzing data] Interpret results [technology and policy for making knowledge persistent and sharable] Deliver tailored message [mechanisms for tailoring messages to decision-makers] Take action to change practice [mechanisms for capturing changed practice]		with sharable and interchangeable components
Leung 2019 [34]	Collecting data from different sources Integrating data from different sources	Clinician	–	Clinical practice guidelines	–	Data represented in a machine processable	Use unique and persistent identifiers,	–	–	A next generation

	Analyzing data from different sources					language	including standard identifiers for key components (e.g. ICD, codes, RxNorm). Version control with an identifier generated through systematic assignment Rich and standardized metadata (released with clear and standard licences), including provenance of how and when they came to be. This should always be available, searchable and accessible.			repository that enables easier access to knowledge and digital objects.
McGraw 2021 [35]	From a reference [1]: A system that produces constantly updated reference data during the care process.	–	–	–	Data generated outside traditional health care: Consumer health and wellness generated data (fitness tracking, medical wearables, medical	–	–	–	Protections for health-relevant data must go beyond pure privacy focus and extend to preventing or penalizing	–

or health monitoring apps, patient-reported outcome surveys, direct-to-consumer tests [including DNA tests] and treatments) Digital exhaust generated (social media posts, internet search histories, location and proximity data) Non-health demographics, social and economic sources (race, gender, income, credit history, employment status, education level, zip code, housing status, census records, bankruptcy and financial records, store purchases, fitness club memberships, voter registration.	uses that could harm individuals and populations.

Platt 2020 [36]	From a reference [21, 22]: Learning (transfer of knowledge through formal curriculae and transfer of culture, attitudes, and believes so it can be implemented in research and practice when combined with quality improvement [21,22] From a reference [1, 25]: Real-time delivity to determine healthcare decisions [1,25] Themes: Culture Innovations Data infrastructure Ethical considerations	Stakeholders in research, clinical, and technical areas	Research, clinical, and technical areas forming interrelated and linked activities.	–	From a reference [26]: Data originates from: clinical practice, research, participation, inquiry provided by rganizational leaders, physicians, researcher, patients, research participant Electronic health records	Mechanisms are in place to ensure that data is usable by all entities and are transferred appropriately, safely, and ethically	–	–	–	From a reference [6]: Data is maintained in a repository until needed for a particular purpose
Vandvik 2020 [17]	Assemble Analyze Interpret Feedback Change	Health care providers Hospital department Patients Health care quality and research agency	Embedding new evidence in daily clinical practice (by using implementation strategies, such as clinical decision support systems) Clinicians document most parts of the care process in a structured fashion in electronic health records. Large repositories of structured data can be fed in to the loop to complete the cycle Integrate research	Patient data Electronic health records Clinical decision support system	–	Make evidence-based offerings more findable, accessible, interoperable, reusable, computable, and useful	–	–	–	Infrastructure to enable a continuous cycle of health improvement, available and secure patient data for research, a knowledge

practices into health care	data-base for health care provid-ers / services / organi-zations to inform health deci-sions.

LITERATURE LIST

1. Akl EA, Haddaway NR, Rada G, Lotfi T. Future of Evidence Ecosystem Series: Evidence synthesis 2.0: when systematic, scoping, rapid, living, and overviews of reviews come together. J Clin Epidemiol. 2020;123:162-5.
2. Bogaert P, van Oers H, Van Oyen H. Towards a sustainable EU health information system infrastructure: A consensus driven approach. Health Policy. 2018;122(12):1340-7.
3. Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. Ann Intern Med. 2020;173(12):1015-7.
4. Cartabellotta A, Tilson JK. The ecosystem of evidence cannot thrive without efficiency of knowledge generation, synthesis, and translation. J Clin Epidemiol. 2019;110:90-5.
5. 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. J Clin Epidemiol. 2020;123:143-52.
6. Dumitrescu A, Granados A, Wallace J, Watson S. Demand-driven evidence network in Europe. Bull World Health Organ. 2006;84(1):2.
7. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11(2):e1001603.
8. Funk T, Sharma T, Chapman E, Kuchenmüller T. Translating health information into policy-making: A pragmatic framework. Health Policy. 2022;126(1):16-23.
9. Gough D, Thomas J, Oliver S. Clarifying differences between reviews within evidence ecosystems. Syst Rev. 2019;8(1):170.
10. Jordan Z, Lockwood C, Aromataris E, Pilla B, Porritt K, Klugar M, et al. JBI series paper 1: Introducing JBI and the JBI Model of EHBC. J Clin Epidemiol. 2022;150:191-5.
11. Lovestone S. The European medical information framework: A novel ecosystem for sharing healthcare data across Europe. Learn Health Syst. 2020;4(2):e10214.
12. Mathews C, Goga A, Loveday M, Zembe W, Daviaud E, Siegfried N, et al. Moving towards universal health coverage: Strengthening the evidence ecosystem for the South African health system. S Afr Med J. 2019;109(11b):8-14.
13. Pilla B, Jordan Z, Christian R, Kynoch K, McInerney P, Cooper K, et al. JBI series paper 4: the role of collaborative evidence networks in promoting and supporting evidence-based health care globally: reflections from 25 years across 38 countries. J Clin Epidemiol. 2022;150:210-5.
14. 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. J Clin Epidemiol. 2020;123:153-61.
15. Shepherd JP. How to achieve more effective services: the evidence ecosystem. 2014.
16. Stewart R, Dayal H, Langer L, Van Rooyen C. The evidence ecosystem in South Africa: growing resilience and institutionalisation of evidence use. Palgrave Communications. 2019;5(1).

17. Vandvik PO, Brandt L. Future of Evidence Ecosystem Series: Evidence ecosystems and learning health systems: why bother? J Clin Epidemiol. 2020;123:166-70.
18. Apfeld JC, Deans KJ. Learning health systems and the future of clinical research. J Pediatr Surg. 2020;55s:51-3.
19. Atkins D, Kilbourne AM, Shulkin D. Moving From Discovery to System-Wide Change: The Role of Research in a Learning Health Care System: Experience from Three Decades of Health Systems Research in the Veterans Health Administration. Annu Rev Public Health. 2017;38:467-87.
20. Britto MT, Fuller SC, Kaplan HC, Kotagal U, Lannon C, Margolis PA, et al. Using a network organisational architecture to support the development of Learning Healthcare Systems. BMJ Qual Saf. 2018;27(11):937-46.
21. de Bruin J, Bos C, Struijs JN, Drewes HW, Baan CA. Conceptualizing learning health systems: A mapping review. Learn Health Syst. 2023;7(1):e10311.
22. Budrionis A, Bellika JG. The Learning Healthcare System: Where are we now? A systematic review. J Biomed Inform. 2016;64:87-92.
23. Casey JD, Courtright KR, Rice TW, Semler MW. What can a learning healthcare system teach us about improving outcomes? Curr Opin Crit Care. 2021;27(5):527-36.
24. Deans KJ, Sabihi S, Forrest CB. Learning health systems. Semin Pediatr Surg. 2018;27(6):375-8.
25. Easterling D, Perry AC, Woodside R, Patel T, Gesell SB. Clarifying the concept of a learning health system for healthcare delivery organizations: Implications from a qualitative analysis of the scientific literature. Learn Health Syst. 2022;6(2):e10287.
26. Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26(2):w107-18.
27. Guise JM, Savitz LA, Friedman CP. Mind the Gap: Putting Evidence into Practice in the Era of Learning Health Systems. J Gen Intern Med. 2018;33(12):2237-9.
28. Institute of Medicine. In: Grossmann C, Powers B, McGinnis JM, editors. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
29. Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. In: Olsen L, Saunders RS, McGinnis JM, editors. Patients Charting the Course: Citizen Engagement and the Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright © 2011, National Academy of Sciences.; 2011.
30. Committee on the Learning Health Care System in America, Institute of Medicine. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
31. Roundtable on Value and Science-Driven Health Care, Forum on Drug Discovery Development and Translation, Board on Health Sciences Policy, Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. Large Simple Trials and Knowledge Generation in a Learning Health

System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
32. A Learning Health System Activity, Roundtable on Value Science-Driven Health Care, Institute of Medicine. Observational Studies in a Learning Health System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
33. Roundtable on Translating Genomic-Based Research for Health, Board on Health Sciences Policy, Institute of Medicine. The National Academies Collection: Reports funded by National Institutes of Health. Genomics-Enabled Learning Health Care Systems: Gathering and Using Genomic Information to Improve Patient Care and Research: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2015 by the National Academy of Sciences. All rights reserved.; 2015.
34. Leung TI, Dumontier M. FAIR Principles for Clinical Practice Guidelines in a Learning Health System. Stud Health Technol Inform. 2019;264:1690-1.
35. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Digit Med. 2021;4(1):2.
36. Platt JE, Raj M, Wienroth M. An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review. J Med Internet Res. 2020;22(3):e17026.

Table S3 - Emerging themes within characteristics of an EES and LHS combined

Characteristic	Theme	Data item
Components, traits, and properties of the system	*Methodological research*	Research on evidence synthesis methods
	Data structuring and storage	Evidence is organized in collections
		Clusters of PICO questions (defined by a common factor)
		Links in a cluster of PICO questions (connecting PICO elements)
		Digitally structured data
		Linked data repositories
		Storage (publication, open access, quality control)
	Generation of data	Generating health data and indicators
		Production of primary research
		Evidence generation (research, expertise, discourse) / generation of evidence / produce evidence
	Syntheses of data	Synthesis of evidence / evidence synthesis (systematic review, evidence summaries, guidelines) / synthesize evidence / production of reviews of primary research / analyze / assemble / synthesize evidence
		Continuous refreshing of evidence syntheses
	Data translation and interpretation	Translating health information
		Translation of evidence (integration of evidence into clinical practice) / health information translation
		Other information used to interpret research findingsEvidence implementation (facilitation of change, evaluation of process and outcome)
		Knowledge translation
		Interpret
	Data curation, transfer, and exchange	Managing health information
		Exchanging health information
		Evidence transfer (active dissemination, systems integration, education)
		Data harmonization via a common data model
		A platform to find, access, and (re)use health data
		Knowledge management (acquire, develop, exchange, disseminate)
		Culture for sharing and innovation

	Tools and platforms
Evaluation and improvement cycle	Policy cycle (agenda setting, policy formulation, policy adoption, policy implementation, policy monitoring and evaluation)
	Use of research for undertaking analysis of situations and informed decisions
	Evaluate and improve policy and clinical practice
	Feedback
	Provide continuous evaluation and improvement
	Embed services and research in a continuous learning loop
	Creating a culture of continuous improvement
Products	Primary research and big data
	Living systematic reviews
	Living evidence services
	Living guidance
	Decision support systems
	Complexity in evidence types
Setting	Health practice
Context	Broader socio-political context
	Global health (sustainable impact, engagement knowledge need)
	Complexity of sectors involved in evidence-informed decision-making
	Foster a sociotechnical perspective, focused on the population
	Anchor in an ultra-large-scale systems approach
	Aligning science and informatics
Actors	Living evidence communities
	Evaluation funders
	Evidence userOrganizations (key organizations in the ecosystem)
	Complexity in individuals and organizations within the system
Governance	Overarching principles (culture, capacity, communication, collaboration)
	Governance to comply to European (privacy) laws and data access

	Common methods in an ecosystem
	Structural foundations (political and constitutional drivers)
	Coordinated processes in an ecosystem
	Global standards
	Orchestration, governance, and support
	Emphasize decentralization and specifications parsimony
	Keep use barriers low and complexity incremental
Activities	Convening (developing meaningful connections between groups, foster consensus, capacity building through shared experience) / Weave a strong and secure trust fabric among stakeholders
	Amplification and advocacy (place issues on global agenda, amplifies member voices, enhances status and legitimacy of individuals in the network)
	Form collaborations (membership, benefits, governance, organizational arrangements, stewardship, resources) Resource mobilization (mobilization for long, medium and short-term initiatives)
	Optimizing trade-offs between speed and thoroughness
	Tailoring end-products to the stakeholder needs
	Engagement of users and producers of research
	Create trustworthy guidance
	Disseminate to policymakers
	Disseminate to clinicians
	Disseminate to patients
	Implement evidence and guidance
	Complexity in the activity spectrum to use activities
	Investments
	Innovations
	Build a shared learning environment
	Leverage existing programs and policies
	Developing strong patient-clinician partnerships
	Providing incentives for innovation
Collaboration	Community building (shared vision, supportive relations)

		Encouraging communities to collaborate and co-create evidence syntheses
		Engage health and healthcare, population, and patient
	Priority-setting and distribution	Priority-setting and global distribution of work
		Complexity in the decision-space for which evidence is sought (range of questions)
		Capacities
	Flow of evidence	Evidence pumps
		Evidence synthesis and guideline production pumps
		Push forces (more prominent in evidence generation)
		Pull forces (more prominent in evidence users)
Actors	*User*	User / end user / evidence user / non-academic user
		Consumers
	Evidence production and synthetization	Evidence synthesizer / authors of systematic reviews / systematic review teams / reviewers
		Producers of information / knowledge producers / evidence producers
		Guideline developers
		Researchers / trialists / academics
		Health technology assessment agencies
		Ethical committees
		Research councils / statutory science councils
		Technology strategy board
		Scientific community
		Research centers
	The public	Citizens
		Communities / Community members / Community centers
		Community-based organizations
		The public
		Non-governmental organizations
	Clinical professionals	Professionals / health professionals / healthcare providers / physicians / clinicians / practitioner-academics / general practitioners / providers

	Accountable care organization
	Hospitals / Hospital groups, hospital systems / Academic health centers
Professional associations	Professional bodies / medical specialty societies / specialist communities / specialty societies
Patient	Patients / patient representatives / patient advocates
	Family members
	Patient associations / patient organizations / patient groups
Resources	Funder / research funder / funding agencies
	Payers / Payors
	CharitiesFoundations
Support	Statisticians
	Methodologists
	Campaigners
	Disseminators
	Data-custodians / Data-owners
	Commissioners
	Planners
	Technology experts
Regulatory	Regulatory agencies / regulatory authorities / regulators
	Local health authorities
Learning	Academia institutions / universities / higher education institutions
	Learning networks
Business	Industry / pharmaceutical companies
	Business and commerce
Policy	Policy-makers
	Governments / Governmental departments
	Health policy authorities
	Politicians
	Decision-makers

		Health care managers / health system managers and leaders / program managers / clinical program leaders / thought leaders (on continuous improvement) / health service managers
		Healthcare administrator
	Dissemination	Biomedical journals / editors / journals
	Media	Media
Links and cooperation	*From networks to any*	Domain specific networks respond to current priorities and carry out analyses
		Domain specific networks can cooperate with national networks for data-collection and data-analysis
		Experts might be commissioned by the network's members to synthesize the best available evidence
		Bringing together networks (national and specific domain) within and between EU member states
	From research to any	Trialists can post protocols in registries
		Trialists can share individual patient data
		Trialists are contacted for missing data
		Researchers offering evidence
		Researchers need approval of data-owners
		Links between primary evidence and evidence synthesis
		Academic centers and departments focus on the production of research to inform achievements of policy, inform decision-making, understanding of the policy, and collecting evidence for decision- making / clinical researchers and clinical program leaders collaborate to establish the evidence through common research priorities and data collection strategies, and use research to inform of progressive iterations of new clinical programs when there is not enough evidence
		Integrate research practices into health care
	From evidence synthesis to any	Definitive findings from well-designed clinical trials shape practice through their incorporation into guidelines, quality measures, or other system-wide implementation efforts
		Evidence synthesis informs the generation of new evidence
		Match between available publications and reports of one trial
		Contact with trialists
		Data within a systematic review can be used for integration with clinical guideline development and clinical decision support systems
		Systematic review authors may bring in topic specialists and users
		Supply-driven reviews

	Demand-driven reviews (engage directly with users)
	Feedback loops between living evidence synthesis production and primary research communities
	Living mapping of evidence to find gaps and support researchers planning a project, funders, and institutional boards
	Evidence syntheses are linked to guidelines
	Producers of systematic reviews tailor the review process and presentation of results to facilitate bringing evidence to practice
	Authors of systematic reviews (methodologists) may bring in other stakeholders (e.g. topic specialists, users, broader societal perspectives) for advise
From evidence translators to any	Evidence translation informs evidence generation
	Evidence translation informs evidence synthesis
	Interpreting research evidence in order to make recommendations or guidance for policy, practice, or individual decision-making
	Guideline developers find data from living meta-analyses on a marketplace and produce locally relevant guide-lines
	Implementing living guidelines and living network meta-analyses helps the research community streamline future clinical trials
	Guideline developers work closely with funding bodies to support evidence generation
	Integration of a guideline developing unit in a specialist community provides an incentive to engage with and apply evidence
From policy to any	Decision-makers can search in a selection of validated web-based sources
	Links between policymakers and those with access to or producers of health information
	Decision-makers seeking evidence
	Close working relations between policy-makers and researchers (e.g. to commission research: demand-driven research, policy-maker pull)
	Researchers are responding to national priorities and strategies
	Government and the research community work together
From point of care to any	For clinicians: journals in hard-copy aimed at their specialties, group discussion, and commissioning teams to summarize evidence
	Clinicians document most parts of the care process in a structured fashion in electronic health records
	Clinician-patient interaction

	Interaction between clinical practice and research domain
	Policy changes and clinical innovations serve to generate evidence via natural experiments that can be used to test specific strategies
	Patients linked with actors producing evidence upstream or those implementing evidence and evaluating impact in practice downstream
	Clinicians, patients, method experts co-create guidelines coordinated with systematic review teams
From commissioners to any	Evidence and cost-benefit need to be applied in the context of such setting, such evidence is generated by action research and should be commissioned (or sought) by commissioners
From unclear to point of care	Uptake of evidence at the point of care (by active dissemination, systems integration, education)
From multiple to multiple	Relationships and interactions form the substance of the network / Key players interact within a national network
	Critical relationship between research and policy or practice and patients
	Integrating research evidence, complexity, perspectives, and norms and values of decision-makers and stakeholders
	Communication, collaboration, and integration of evidence between creators and end-users
	Cooperation between key stakeholders
	Create relations between evidence synthesis community and guideline developers, regulators, and governments to support, promote and sustain living evidence synthesis
	Link between living evidence synthesis community, researchers, decision-makers for specific clinical questions and living network meta-analyses
	Patients, clinicians, family members, health system leaders are working together to coproduce healthcare
	Patients, clinicians, family members, healthcare leaders working together as partners
	Research, clinical, and technical areas forming interrelated and linked activities
	Interaction between stakeholders (joint goals-setting, revealing underlying values, evaluation processes, share best practices)
	Clinicians, patients, researchers, community members, community-based organizations, payors and other stakeholders are actively engaged in the learning process and/or clinical decision-making and/or improving the organization and/or research
Governance	Stakeholders (patients, families, clinicians, researchers) participate in all aspects of governance
	A centralized network built around a governing entity with nested groups organized by region (structurally managed with regional chairs to foster regional collaborative efforts [e.g. mentorship, support of novice groups])
Flow and direction	The flow through the ecosystem is not unidirectional

	Local circumstances may influence the entrypoint into the flow (e.g. feasibility, burden of disease)
	Bidirectional relations between global health, evidence generation, evidence synthesis, evidence transfer, evidence implementation)
	Mutual exchange of ideas between policy-makers, practice, research, personal decisions,
	Up to date (global and local) evidence needs to flow efficiently between producers, synthesizers, processors and implementors
	Bidirectional interactions with stakeholders
	A system needs conduits connecting each stage to transport from storage facilities to distribution points, with interfaces for end-users)Institutions and roles which provide more than one ecosystem function help to connect the entire system
	Evidence needs to flow through the ecosystem from generation to end-user
	Organizations driving both the supply and demand of evidence
	Co-production and networking bring internal and external initiatives closer
	Data can flow between interoperable platforms for systematic reviews, guidelines, and clinical decision support systems in electronic health records
Within evidence	Intervention and control (within a cluster)
	Links are populated with all relevant studies and a standardized data file
	Product blending (e.g. evidence from process evaluations may be very useful blended with data about effectiveness)
Activities	Cocreation to help tailored outputs
	Evidence needs to be synthesized and translated into policy and guidelines
	Downstream decision support could be linked to active implementation of evidence into practice and health care systems
	Embedding new evidence in daily clinical practice (by using implementation strategies and clinical support systems)
	Distributing workload for evidence syntheses among teams at the international levelResearch prioritization involving stakeholders
	Stakeholder engagement
	Experienced and agile teams of software engineers, database engineers, extract transform and load experts, computer scientists, and analysts create the infrastructure
Qualities of activities	Commitment of clinicians, quality improvement specialists, healthcare system leadership, researchers, and patients

		Relationships are multifaceted, ranging from structured, formal relationships with goals and objectives (e.g. collaboration agreements, mentorships, clinical partnerships) to social relationships
		Active collaboration from all partners in the system
	Repositories and platforms	Large repositories of structured data can be fed into the loop
		Training resources for multiple stakeholders
		Distributed structure with a central hub that coordinates the operation of distributed facilities or networks
Products	*Questions*	PICO question
	Evidence syntheses	(Network)meta-analysis / evidence synthesis / evidence reports / (Living) systematic review
		Qualitative comparative analysis
		Reviews of reviews
		Scoping review
		Rapid review
		Text and opinion evidence synthesis
		Prevalence and incidence evidence synthesis
		Economic evidence synthesis
		Etiology and risk evidence synthesis
		Mixed methods evidence synthesis
		Umbrella reviews
		Measurement properties reviews
		Health technology assessment appraisals (costs, effectiveness, impact of health care)
		Performance expenditure reviews
		Responsive evidence synthesis
	Evidence summaries	Evidence summaries / evidence summaries from reviews / summary results
		Evidence brief for policy
	Guidelines	Clinical practice guidelines / living guidelines / clinical guidelines and related derivative content / digitally structured guidelines
		Standards
	Evidence maps	Evidence map / maps / gap maps / living mapping of research / evidence (gap) map / Policy-relevant evidence mapping

	Knowledge gapsGaps in primary research
	Living map of all trials
	Maps of maps
Notifications	Local health messages
	Memoranda
Research output	Research reports / clinical study report
	Public health reports / annual reports
	Prediction models / prognostic and predictive models
Type of research	Primary health systems research
	Primary research
	Observational research
	Multi-site or single-site RCTs / trial data / randomized controlled trials / comparative effectiveness research
	(Large) pragmatic trials
	Quasi-experiments
	Impact analyses
Data	Real-world data / routinely collected data / population-based data in registries / electronic health record data / patient data / health information / care plans
	High-level meta-data
	Observational data / large longitudinal datasets / research cohort dataQualitative information / qualitative data from RCTs for effectiveness
	Quantitative evidence for stakeholder views, acceptability, and feasibility
	Participants and program staff anecdotes and observations
	Monitoring data / Citizen monitoring data
	Administrative data
	Analysis-ready dataset
	Individual patient data
	Economic evidence for resource use
Decision aid	HTA decision aids

		Interactive decision aids / clinical decision support systems / decision support systems
		Data-dashboards
	Support	Protocols
		Case report forms
		Statistical analysis forms
		Trial registrations
		Quality indicators
		Health plans
	Policy	PoliciesLetter to or from government
Data sources (LHS only)	*Location*	Clinical practice / Data collection in clinical practice
		Data flowing from routine care
	Actors	Physicians
		Researcher
		Patients
		Research participant
	Products	Electronic health records / Personal health records / Clinical data
		Use-case cohorts
		Research studies / biomedical research
		Electronically captured and combined data from patient-clinician interaction
		External evidence from trials, studies, and reviews / Randomized controlled trials / Observational studies / Systematic review / Medical product research
		Clinical dashboards
	Data types	Claim information
		Lab samples
		Genomic data
		Consumer health and wellness generated data (fitness tracking, medical wearables, medical or health monitoring apps, patient reported outcome surveys [including DNA tests and treatments])

		Non-health demographics, social, and economic sources (race, gender, income, credit, history, employment status, education level, zip code, housing status, census records, bankruptcy and financial records, store purchases, fitness club memberships, voter registration.
		Clinical and administrative data / Clinical history
		Health related data
		Social determinants of health
		Patient generated data
		Patient reported measures or outcomes
		Experience of care (patient and professional) / Annual survey on patient experience
	Database characteristics	Computer searchable databases
		Distributed but networked databases
	Databases	Clinical trial databases
		Biobanks
		Health information databases
		Registries
		Corporate warehouse
		Claims databases
		Pharmaceutical clearinghouses
	Used for data generation	Research
		Participation
		Inquiry provided by organizational leaders
		Telehealth
		Electronic monitoring devices
		Mobile sensors
Data presentation / packaging	*Easy to interpret*	Interactive data visualization / data-visualization / visually appealing outputs
		Interactive infographics with direct links to multi-layered digital content
		Interactive decision aids
	Products	Descriptive data / Text, tables, figures

	Forest plots
	GRADE summary of findings tables
	Evidence profiles / Risk of bias tables
	Logic model to describe and document theories of change
Tools	Information packaging and synthesis toolsCommunication and dissemination tools
	Information linkage and exchange tools
	Visualization and modeling tools (Graphs, charts, maps, modeling, and simulation)
Interoperability	A database taxonomy based on meta-data (including a data-dictionary)
	common data model (for harmonization, to standardize the format and content of observational data)
	Standards for reporting data (to bypass data-extraction and include data straight into syntheses) /
	Standardized data file (containing study characteristics, risk of bias assessment, summary results, individual patient data)
	Comprehensive data models and semantic interoperability mechanisms / Computable representation / data represented in machine processable language
	Structured publication and verification of research output using semantic technologies
	Translation mechanisms (push, pull, exchange, integrated)
	Linkage and/or storage of different data-sources
Packaging	Packaging review findings and guidelines into user-friendly products
	Products should have various formats, sizes, different focus of context, different languagecomplexities, and translated to different languages
	Evidence categorization according to the practitioners' and commissioners' need for usable forms
	Digital, multi-layered and user-friendly reports
	Promoting evidence syntheses in short summaries / evidence reports (i.e. syntheses of best available evidence in an easy-to-read style)
	Standardized representation
	Fit for purpose and easy to understand
Accessibility	Guidelines and decision aids on all devices and in web-portals
	Ready-for use, adaptations, and plug-ins
	Mechanisms are in place to ensure data is usable by all entities and are transferred appropriately, safely, and ethically

		Meta-data to allow sub-grouping and filtering according to user needs
		Search engine presenting relevant evidence for the PICO
Values and procedures for quality	*Reproducibility*	Reporting guideline adherence (e.g. CONSORT)
		Reproducibility of research and clinical practice guidelines (e.g. EQUATOR network, AGREE)
		Standards of synthesis (methods, included studies, totality of evidence produced)
		Reporting standards
		Platforms that incorporate trustworthy standards for decision support systems
		Timely, clear, usable, relevant, credible evidence generated through uncontested methods
		Generation of evidence using appropriate methods
		Standard operating procedures
		Cochrane risk of bias tool / risk of bias assessments
		In-house minimal dataset for cohort data
	Products and frameworks	Core outcome sets
		GRADE approach
		Knowledge translation frameworks
		GRADE Evidence-to-decision framework
		Standard framework for quality improvement
	Meta-data	Rich and standardized meta-data which should always be available, searchable, and accessible.
		Rich meta-data
	Data sharing	Unique and persistent identifiers including standard identifiers for key components / version control with identifiers
		Intended and realized data sharing / data sharing
		Promoting best practices in the production, updating and dissemination of systematic reviews and guidelines (e.g. handbooks and through network [e.g. GIN])
		Making evidence-based products more findable, accessible, reusable, computable and useful / curation of research to make data easy to find, accessible, and reusable / Ensuring access to quality information
		Living disclosure of the quality, transparency and accessibility of data
		Sharing of trial data

	Collaboration (government will work effectively across agencies, with state and local governments, and with others outside the government)
Processes and actions	Conflict of interest disclosure and management
	Identifying high priority areas (e.g. James Lind)
	Identifying existing and potential problems, and addressing those problems systematically
	Use of new sources for data (e.g. registries, journal appendices, reporting protocols, regulatory agencies, pharmaceutical companies)
	Crowd sourcing (citizen science)
	Guidance about types of evidence syntheses
	Skills training for research assistants in commissioning teams
	Building a safety culture
	Establish a process of gradual improvement / Improving quality
	Controlling costs
	Public participation (welcoming an accessing widely dispersed national interests and expertise to help the government improve the well-being of the people)
	Coordinating complex work across diverse organizations
Transparency	Transparency (publishing data with the intent of improving government accountability and enabling citizens to access and use data in ways that are socially beneficial)
	Transparency of process, outcome, price, and cost information
	Open government
	Transparency by clinical trial and systematic review registration
Privacy and regulation	Commitment that data will not be used for competitive purposes or be put on public display
	National electronic health record data, registry data, connectivity, reporting, and privacy protection standards
	Standardized processes, protocols, and policies (communication, data-sharing, privacy protection, regulatory compliance)
	Careful and constant attention to privacy and security issues
Waste and publication bias	Reduce research waste (e.g. REWARD alliance)
	Sharing of review data to prevent duplication
	Recognize and minimize research waste, unusable evidence, and contaminated evidence (e.g. wrong methods)
	Post results in trial registry for transparent and complete reporting

		Access to trial protocols
	Technology	Automated technology (assisting evidence synthesis, AI, machine leaning, text mining, data linkage)
		New technology and AI to improve efficiency and reliability of outputs
		Standardized technology to reduce data entry-burden, facilitate care, and which results in useful data for learning and research
Updating/maintenance (EES only)	*Rapid processing*	Rapid recommendations
	Continuous processing	Living systematic reviews (based on living protocol)
		Living network meta-analysis
		Living guidelines
		Living evidence
	Support	Artificial intelligence and automation identifying, coding, and abstraction of data
		A process for matching emerging evidence to the relevant collection or cluster (i.e. a living universe)
Learning and evaluation (LHS only)	*Actions*	Continuous testing of interventions in real-world settingsResearch prioritization involving existing evidence and network outcome/process data
		Generate and test hypotheses
	Sharing	Common repository to share resources
		Shared knowledge
		Shared tools and resources
		Shared standards
		Shared situational awareness
		Sharing ideas, knowledge, and know-how (all teach all learn culture)
		High degree of transparency and ongoing sharing of measures and aggregate and site-level performance
	Mirroring and evaluation	Observe own results and compare them to others' experiences
		Networks evaluate their policies in regular cycles
		Evaluate management structure regularly
		Evaluate the impact of research resources
		Quality improvement through peers changing each other's behavior based on historical performance
		Monthly calls and meeting to evaluate the network functioning

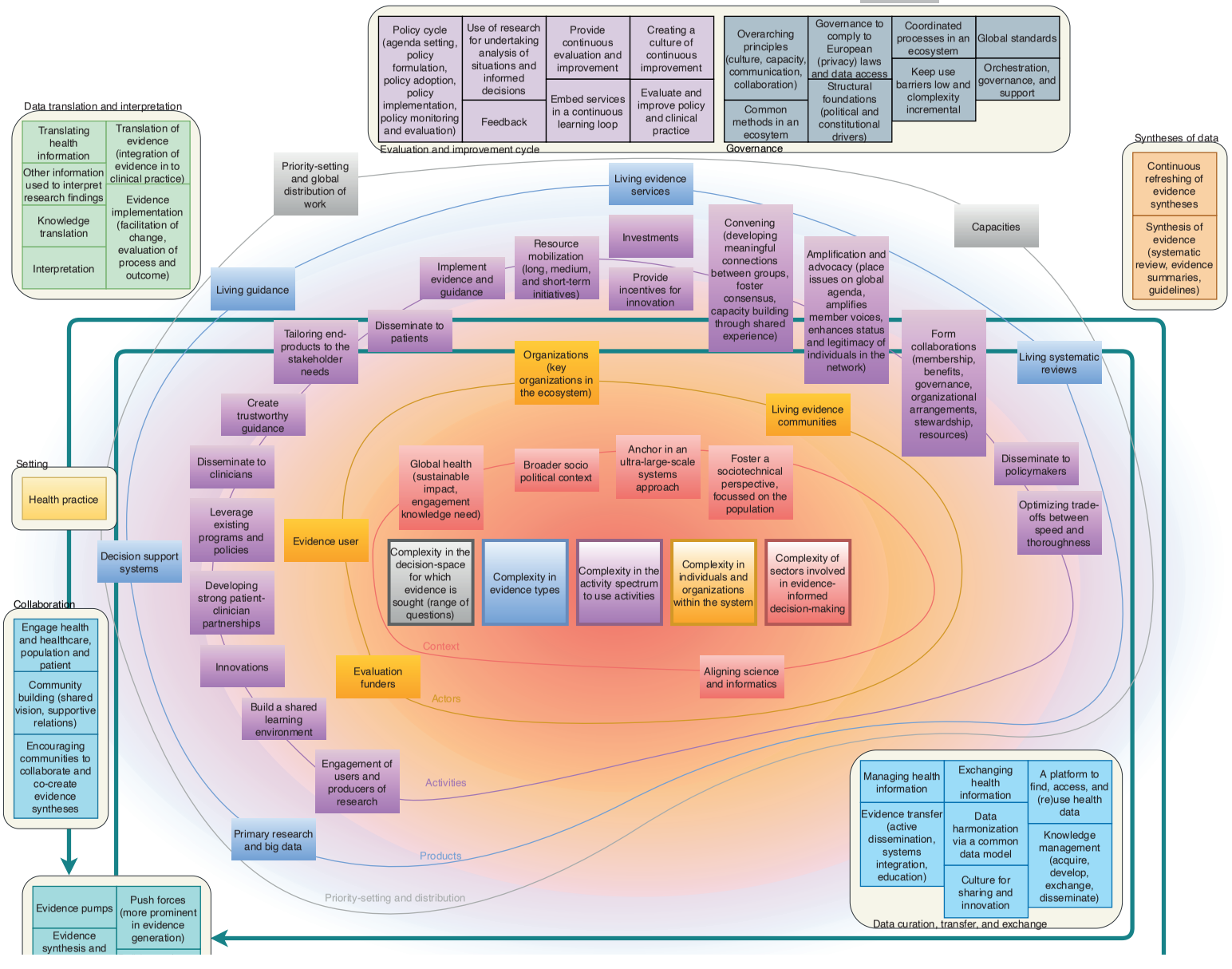
	Evaluate the effects of implementation to redefine or de-implement ineffective or harmful interventions
	Learn from large numbers of patient experiences through electronic health record systems
Learning loops	Point of care, research during normal clinical care, integration in knowledge networks (including biomedical research), application of best available evidence (taking in to account patient needs, local system, community context)
	Performance to data, data to knowledge and adding critically appraised external evidence, knowledge to performance, etcetera.
	Learning cycle: decision to study, assemble relevant data, analyze data, interpret results, deliver tailored message, take action to change practice
	Defining the problem, developing algorithms, implementing tools, evaluating performance of the tools and their effect on patient care and outcomes
	Data and information - analyzed - problematic factors identified - action identified - actions assigned, performed, monitored - feedback to clinicians, employees, leadership - validation - data and information - loop continues
Learning types	Learning as (intermittent) information exchange between the clinical domain and the research domain / Research influences practice, practice influences research
	Learning as a (technology) aided continuous circular process of converting (routine care) data into knowledge, knowledge to performance, performance to data
	Learning as a recurrent interaction between stakeholders to reveal/discuss underlying processes, and to identify opportunities for change and share best-practices
Enablers	Relevant, real-time results to enable more effective care, to encourage participatory and science-based decision-making, and to foster continuous learning for all healthcare stakeholders
	Organizational culture emphasizes and supports learning (transparency, collaboration, culture facilitates trust building)
	Learning is championed by organizational leaders (clinicians encouraged to conduct research, employees as active learners)
	Constantly refining care operations and care processes through ongoing team training, skills building, system analysis, information development, and creation of feedback loops
	Leadership incentives and culture support for systematically capturing and translating information generated by clinical research and care delivery
	Strong academic affiliations
	Clinicians with dual appointments
	Statistical process control methods
	Strong emphasis on systems redesign and learn training

		Collaborative quality improvement (integrating subject and improvement knowledge, ongoing learning, alternate learning and action periods)
Governance	*Levels*	Decision-making level (assembly of members, a scientific advisory board, ethics board, privacy board)
		Executive level (central executive management office, consultation platform)
		Operative level (network committee)
		Organization at a global level
	Types	Local provenance and governance (tools on platforms run locally at the data-custodian's site)
		Governance referring to operational structures, roles, responsibilities, and decision-making
		Governmental monitoring and evaluation system (including citizen monitoring to monitor service quality)
		Network governance of membership policies (guidelines, rights, obligations)
		Network governance of collaboration and attribution policies (authorship, copyright, intellectual property)
		Network governance of data sharing policies (access, ownership, privacy, security)
		Network governance of research and regulatory policies (informed consent, master reliance agreement)
		Network governance of privacy oversight (patient generated data)
		Governance of evidence through an independent quality assurance framework (including a process for declaring conflicts of interest)
		Ethical code of practice for ethical governance (to ensure data protection, patient confidentially, compliance with European privacy laws)
	Boards and committees	Assembly of full members
		Scientific advisory board
		Ethics and privacy board
		Central executive management office
		Consultation platform
		Network committee
		Independent steering committee
	Traits and tasks	Safeguards against undue influence of lobby bodies
		System sensors to identify and correct faults
		Overarching infrastructure to govern actors working across the steps
		Strict management of conflicts of interest

		Strict management of patient partnership
		Protections beyond privacy focus by preventing or penalizing uses of health-relevant data that could harm individuals and populations
		Driving strategic and operating standards for public service administration
		Approval needed to export data to external software
		Full decision making power
		Offer advice
		Management, operational and budgetary day-to-day decisions
		Align with health landscape
		Oversee national scientific activities
		Privacy protection of data subjects and to protect the interests of all data-sharing parties
	Members	Members and observers (in: assembly of full members)
		Scientific experts (in: scientific advisory board / ethics and privacy board)
		General director and core team (in: central executive management office)
		International organizations and commission services (in: consultation platform)
		Representatives of national health information authorities and international research networks (in: network committee)
Platforms	*Platforms*	Web-based platforms / website
		Virtual and interoperable repository platforms / a repository
		Electronic health record search software
	Tools	Evidence mapping software applications for visualizations
		A range of platforms for synthesis (such as Covidence and Cochrane RevMan) / Specialist systematic review software from international organizations (Cochrane, EPPI-center) for the management of large amounts of research
		Platforms to process evidence (GRADEpro, MAGICapp)
		Platform contains a set of tools to enable data-discovery and characterization, documentation, task management, process management and analysis in a secure environment / Tools in the platform were: a catalogue, federated analyses of electronic health records, patient selection tools, task management, private remote research environment, codemapper

	Uses	A marketplace for evidence syntheses
		A secure data facility where administrative data is prepared and curated for research and policy analysis
		A practical repository solution to share trial data /platforms that facilitate data sharing across the evidence ecosystem / Platform to find, assess and (re)use health data from diverse sources across Europe / A repository of all relevant trials classified according to an ontology can be used by organizations to describe the included trials and share data among each other (may include data-extraction, risk of bias assessments, selection of outcomes, interpretation of findings) / Digital libraries to curate and manage digital knowledge objects / a next generation repository that enables access to knowledge and digital objects
		Interoperable platforms for systematic reviews, guidelines, and clinical decision support systems in electronic health records / The platform as an integration platform, where data is harmonized to new data sets and stored on the platform / Support system
		Infrastructure to enable continuous cycles of health improvement, available and secure patient data for research, a knowledge database for healthcare providers, services, informed health decisions
	Architecture	Storage points (with end-user interfaces)
		A central portal of repositories from different universities
		Network of nodes that locally process sensitive patient data and shares products of computationCentralized architecture
		Centralized architecture
		Creating a digital infrastructure by software engineers, database engineers, extract transform load experts, computer scientists, informaticists, analysts.

EES: Evidence Ecosystem
LHS: Learning Healthcare System


Policy cycle (agenda setting, policy formulation, policy adoption, policy implementation, policy monitoring and evaluation)
Use of research for undertaking analysis of situations and informed decisions
Feedback
Provide continuous evaluation and improvement
Embed services in a continuous learning loop
Creating a culture of continuous improvement
Evaluate and improve policy and clinical practice
Evaluation and improvement cycle
Overarching principles (culture, capacity, communication, collaboration)
Common methods in an ecosytem
Governance to comply to European (privacy) laws and data access
Structural foundations (political and constitutional drivers)
Coordinated processes in an ecosystem
Keep use barriers low and clomplexity incremental
Global standards
Orchestration, governance, and support
Governance
Data translation and interpretation
Translating health information
Other information used to interpret research findings
Knowledge translation
Interpretation
Translation of evidence (integration of evidence in to clinical practice)
Evidence implementation (facilitation of change, evaluation of process and outcome)
Syntheses of data
Continuous refreshing of evidence syntheses
Synthesis of evidence (systematic review, evidence summaries, guidelines)
Priority-setting and global distribution of work
Living evidence services
Investments
Provide incentives for innovation
Resource mobilization (long, medium, and short-term initiatives)
Implement evidence and guidance
Convening (developing meaningful connections between groups, foster consensus, capacity building through shared experience)
Amplification and advocacy (place issues on global agenda, amplifies member voices, enhances status and legitimacy of individuals in the network)
Capacities
Living guidance
Tailoring end-products to the stakeholder needs
Disseminate to patients
Organizations (key organizations in the ecosystem)
Form collaborations (membership, benefits, governance, organizational arrangements, stewardship, resources)
Living systematic reviews
Create trustworthy guidance
Living evidence communities
Disseminate to clinicians
Global health (sustainable impact, engagement knowledge need)
Broader socio political context
Anchor in an ultra-large-scale systems approach
Foster a sociotechnical perspective, focussed on the population
Disseminate to policymakers
Setting
Health practice
Leverage existing programs and policies
Evidence user
Optimizing trade-offs between speed and thoroughness
Decision support systems
Complexity in the decision-space for which evidence is sought (range of questions)
Complexity in evidence types
Complexity in the activity spectrum to use activities
Complexity in individuals and organizations within the system
Complexity of sectors involved in evidence-informed decision-making
Developing strong patient-clinician partnerships
Collaboration
Engage health and healthcare, population and patient
Community building (shared vision, supportive relations)
Encouraging communities to collaborate and co-create evidence syntheses
Context
Innovations
Evaluation funders
Aligning science and informatics
Actors
Build a shared learning environment
Engagement of users and producers of research
Activities
Primary research and big data
Products
Priority-setting and distribution
Evidence pumps
Evidence synthesis and
Push forces (more prominent in evidence generation)
Managing health information
Exchanging health information
A platform to find, access, and (re)use health data
Evidence transfer (active dissemination, systems integration, education)
Data harmonization via a common data model
Knowledge management (acquire, develop, exchange, disseminate)
Culture for sharing and innovation
Data curation, transfer, and exchange

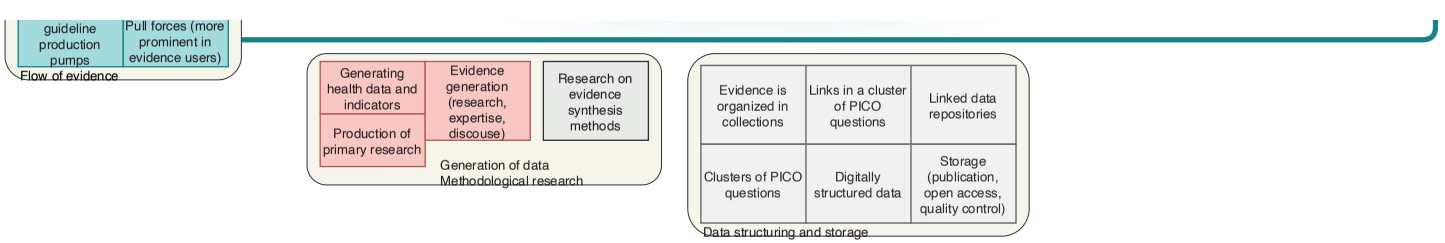


Figure S1. Intermediary visualization of the 'Components, traits and properties of the system' characteristic. *The visualization shows possible associations and/or mechanisms between the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

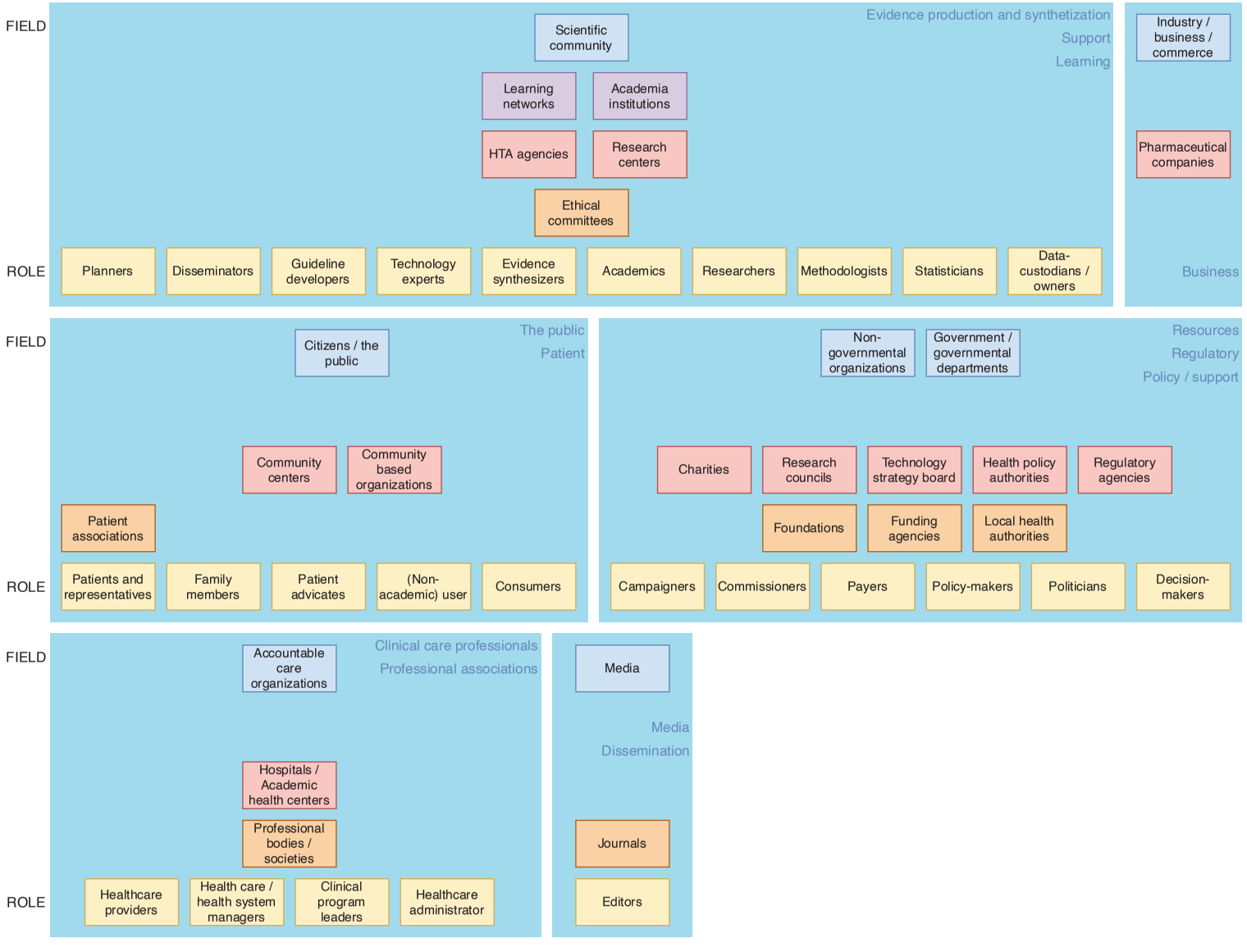

FIELD
ROLE
Scientific community
Learning networks
Academia institutions
HTA agencies
Research centers
Ethical committees
Evidence production and synthetization
Support
Learning
Industry / business / commerce
Pharmaceutical companies
Business
Planners
Disseminators
Guideline developers
Technology experts
Evidence synthesizers
Academics
Researchers
Methodologists
Statisticians
Data-custodians / owners
Citizens / the public
The public
Patient
Community centers
Community based organizations
Patient associations
Patients and representatives
Family members
Patient advicates
(Non-academic) user
Consumers
Non-governmental organizations
Government / governmental departments
Resources
Regulatory
Policy / support
Charities
Research councils
Technology strategy board
Health policy authorities
Regulatory agencies
Foundations
Funding agencies
Local health authorities
Campaigners
Commissioners
Payers
Policy-makers
Politicians
Decision-makers
Accountable care organizations
Clinical care professionals
Professional associations
Hospitals / Academic health centers
Professional bodies / societies
Healthcare providers
Health care / health system managers
Clinical program leaders
Healthcare administrator
Media
Media
Dissemination
Journals
Editors

Figure S2. Intermediary visualization of the 'Actors' characteristic. *The visualization shows an overview of actors and their possible hierarchy within a cluster (blue field) resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

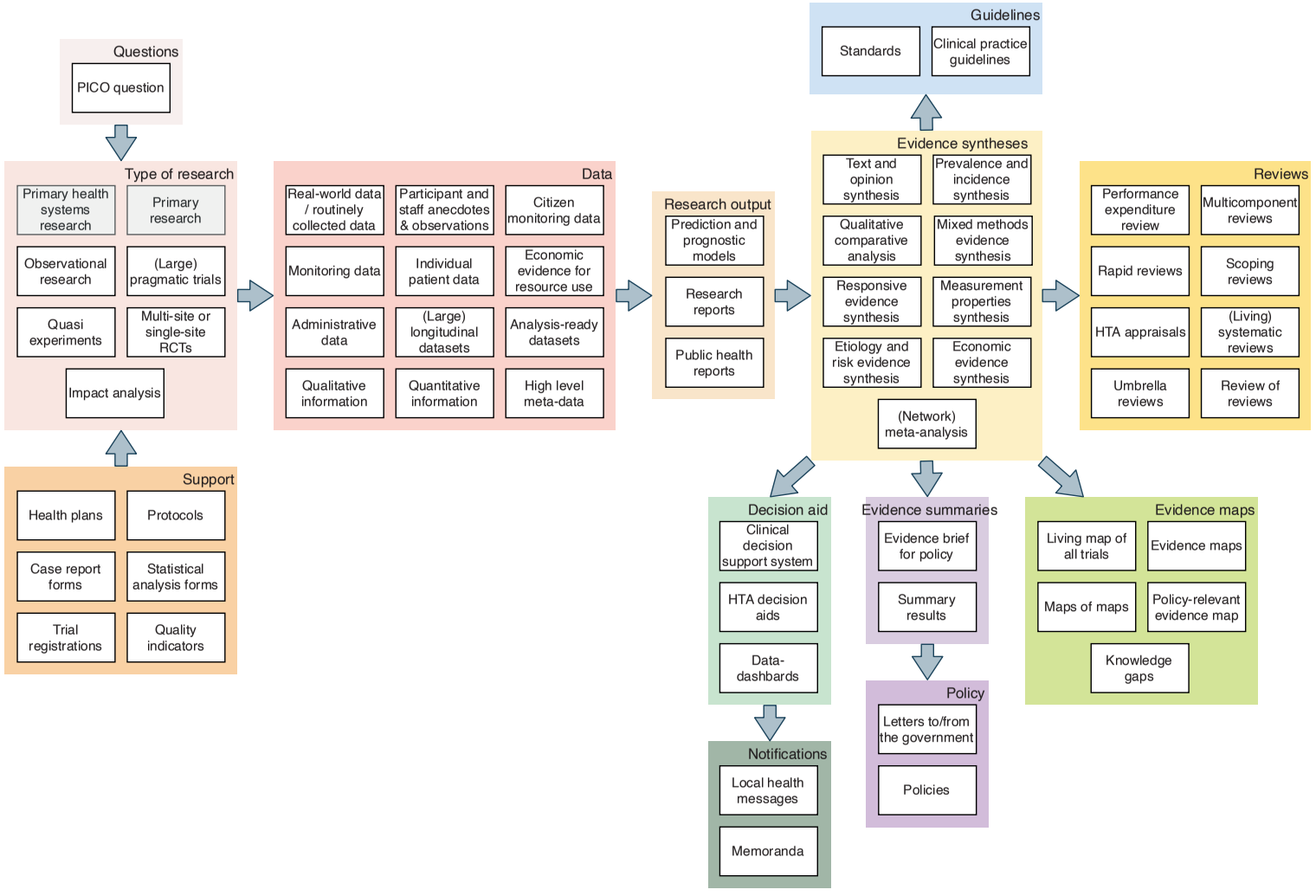

Questions
PICO question
Type of research
Primary health systems research
Primary research
Observational research
(Large) pragmatic trials
Quasi experiments
Multi-site or single-site RCTs
Impact analysis
Support
Health plans
Protocols
Case report forms
Statistical analysis forms
Trial registrations
Quality indicators
Data
Real-world data / routinely collected data
Participant and staff anecdotes & observations
Citizen monitoring data
Monitoring data
Individual patient data
Economic evidence for resource use
Administrative data
(Large) longitudinal datasets
Analysis-ready datasets
Qualitative information
Quantitative information
High level meta-data
Research output
Prediction and prognostic models
Research reports
Public health reports
Guidelines
Standards
Clinical practice guidelines
Evidence syntheses
Text and opinion synthesis
Prevalence and incidence synthesis
Qualitative comparative analysis
Mixed methods evidence synthesis
Responsive evidence synthesis
Measurement properties synthesis
Etiology and risk evidence synthesis
Economic evidence synthesis
(Network) meta-analysis
Reviews
Performance expenditure review
Multicomponent reviews
Rapid reviews
Scoping reviews
HTA appraisals
(Living) systematic reviews
Umbrella reviews
Review of reviews
Decision aid
Clinical decision support system
HTA decision aids
Data-dashbards
Notifications
Local health messages
Memoranda
Evidence summaries
Evidence brief for policy
Summary results
Policy
Letters to/from the government
Policies
Evidence maps
Living map of all trials
Evidence maps
Maps of maps
Policy-relevant evidence map
Knowledge gaps

Figure S3. Intermediary visualization of the 'Products' characteristic. *The visualization shows an overview of products and their possible relations resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

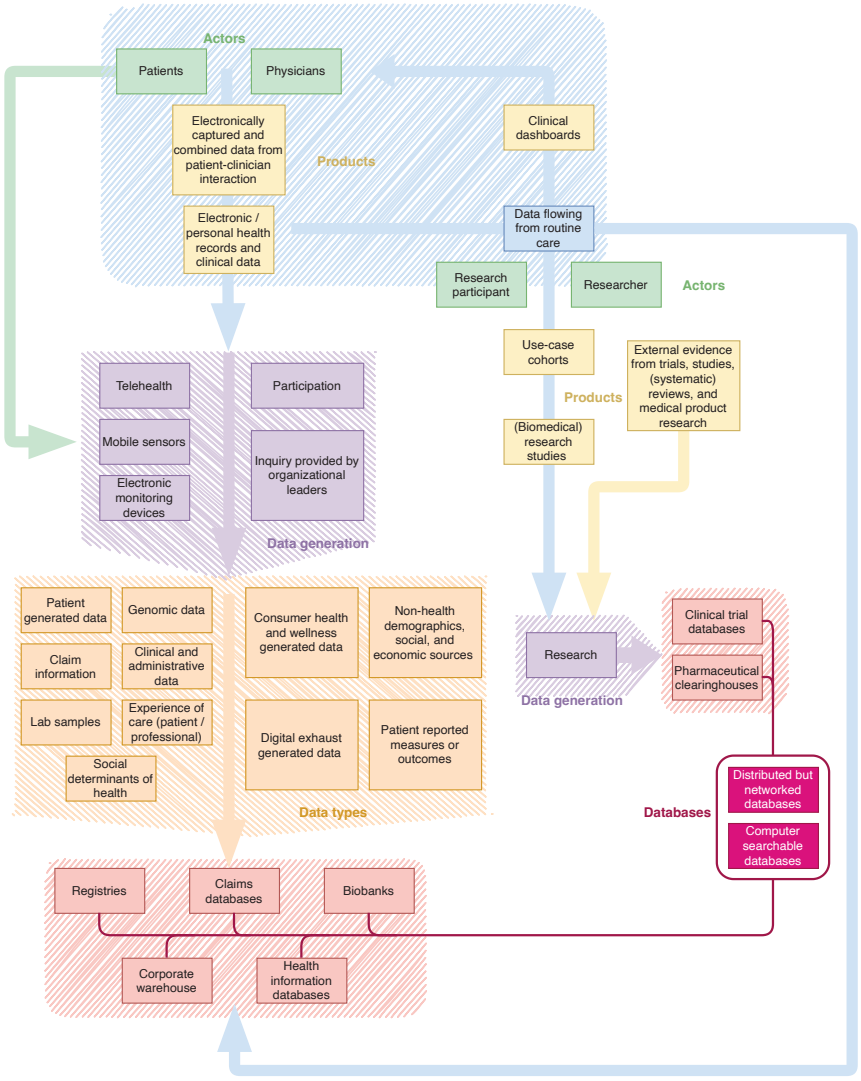


Figure S4. Intermediary visualization of the 'Data sources (LHS only)' characteristic. *The visualization shows possible relationships of actors, data sources, products and databases resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

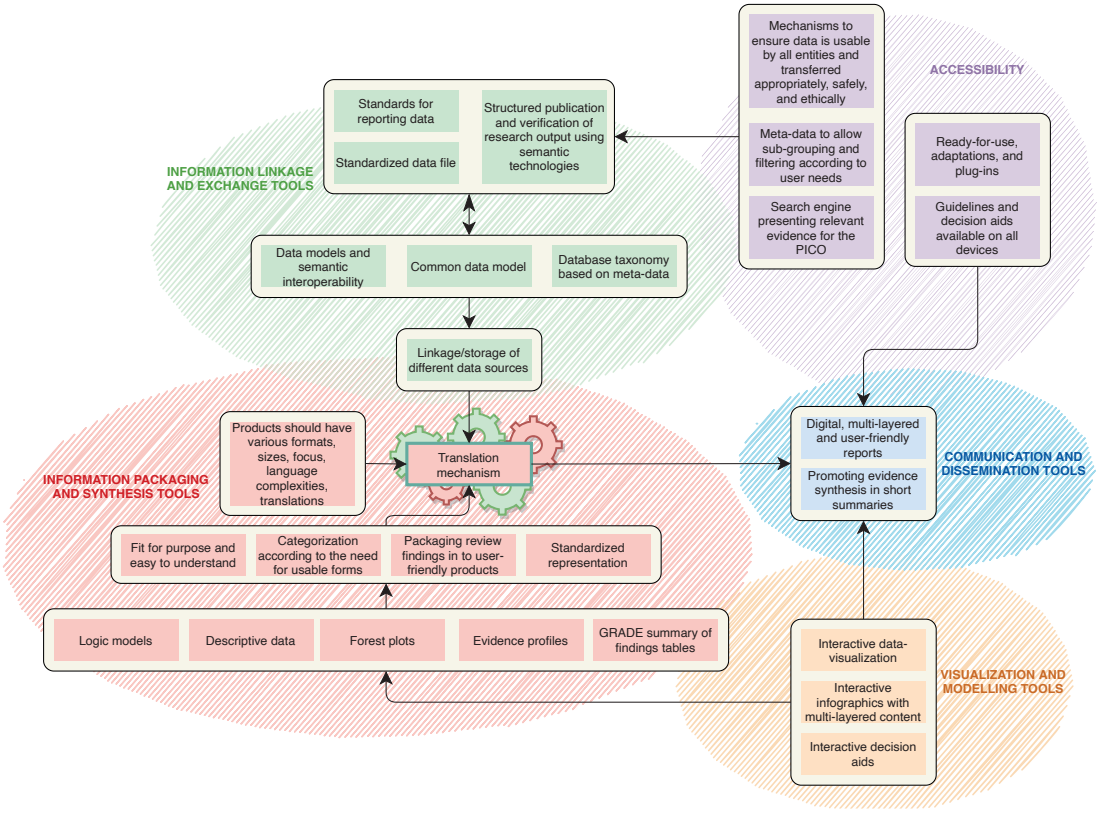


Figure S5. Intermediary visualization of the 'Data presentation/Packaging' characteristic. *The visualization shows an overview of how information is possibly linked, accessed. exchanged, packaged, visualized, and communicated resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

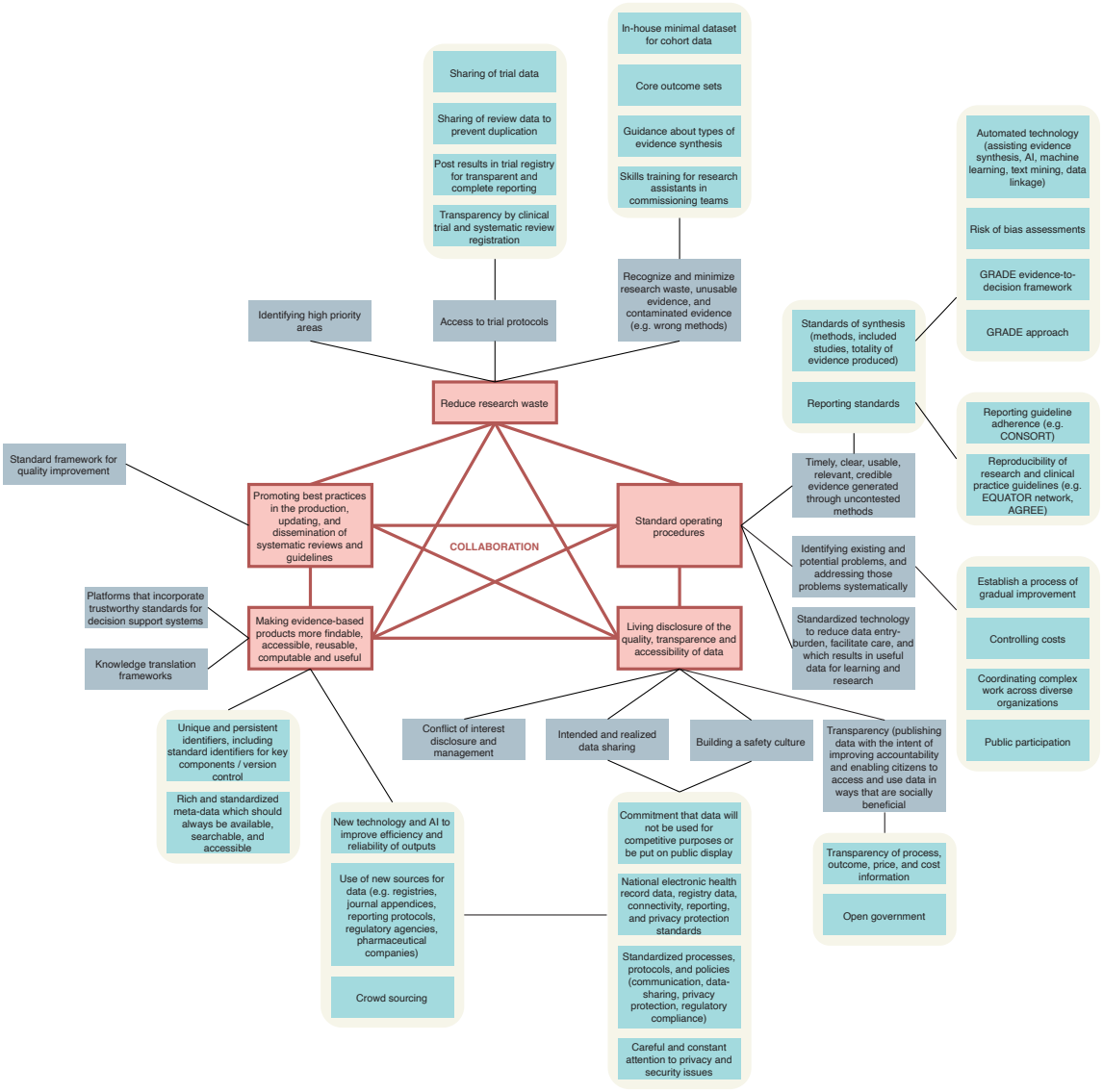


Figure S6. Intermediary visualization of the 'Values and procedures for quality' characteristic. *The visualization shows an overview of possible relations between values and procedures resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

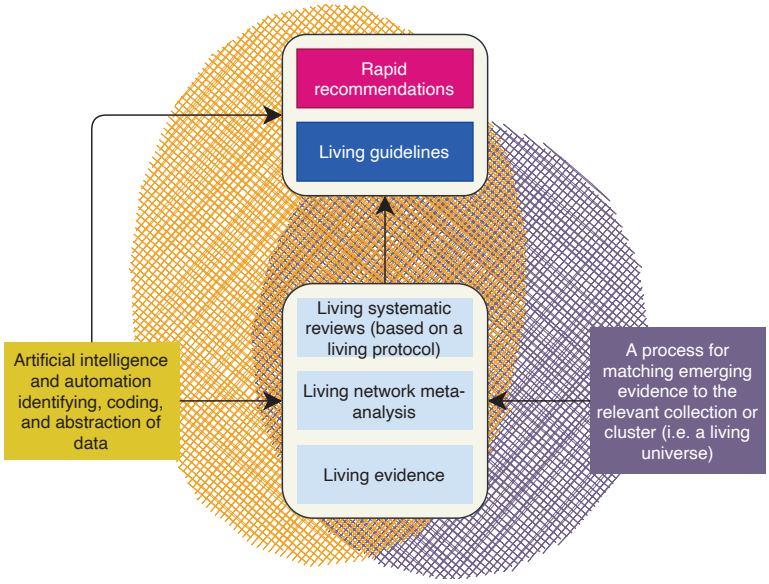


Figure S7. Intermediary visualization of the 'Updating/maintenance (EES only)' characteristic. *The visualization shows a possible process for maintaining products through a living process of evidence for rapid recommendations and/or living guidelines, which resulted from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

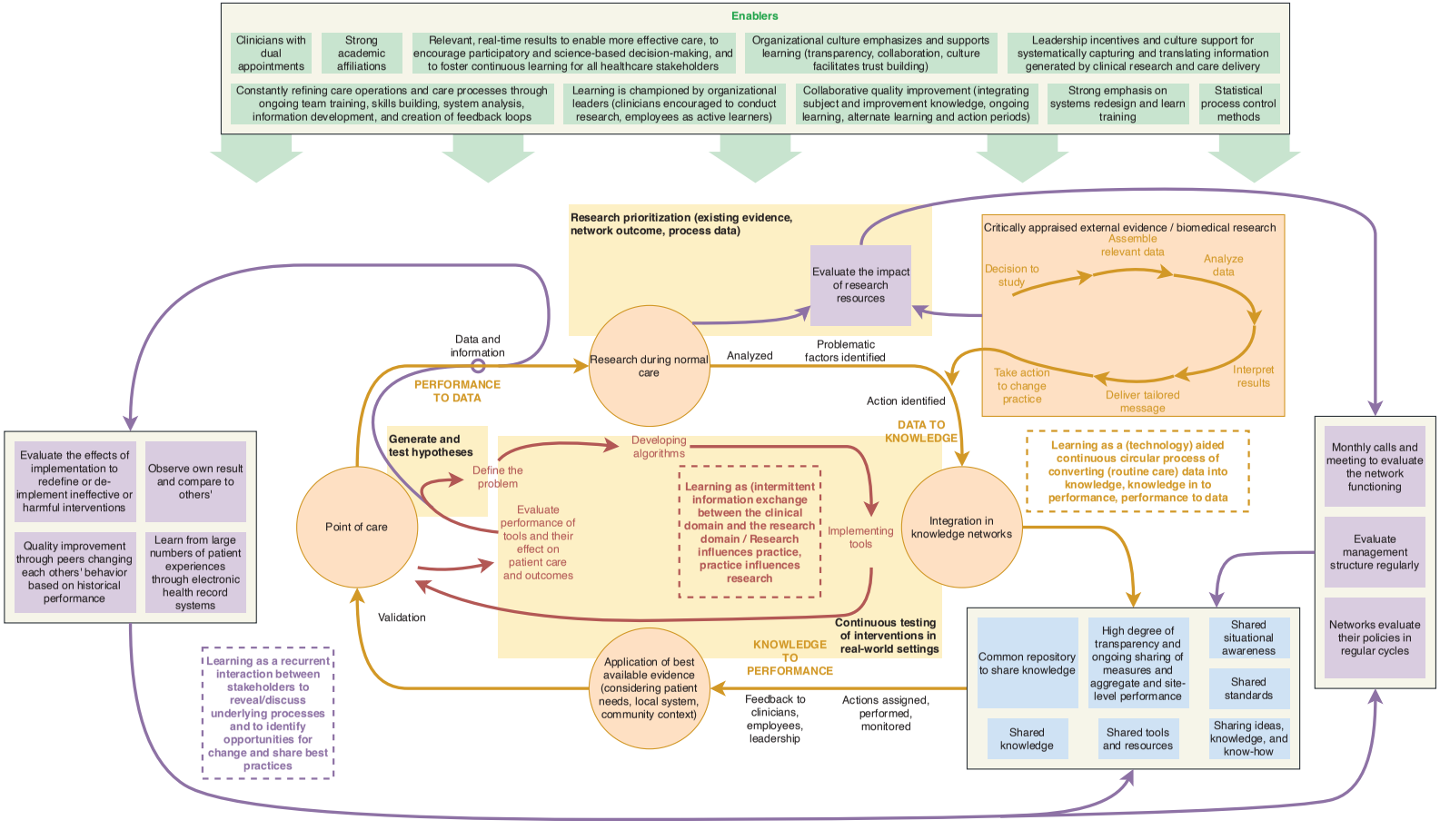

Enablers
Clinicians with dual appointments
Strong academic affiliations
Relevant, real-time results to enable more effective care, to encourage participatory and science-based decision-making, and to foster continuous learning for all healthcare stakeholders
Organizational culture emphasizes and supports learning (transparency, collaboration, culture facilitates trust building)
Leadership incentives and culture support for systematically capturing and translating information generated by clinical research and care delivery
Constantly refining care operations and care processes through ongoing team training, skills building, system analysis, information development, and creation of feedback loops
Learning is championed by organizational leaders (clinicians encouraged to conduct research, employees as active learners)
Collaborative quality improvement (integrating subject and improvement knowledge, ongoing learning, alternate learning and action periods)
Strong emphasis on systems redesign and learn training
Statistical process control methods
Research prioritization (existing evidence, network outcome, process data)
Evaluate the impact of research resources
Critically appraised external evidence / biomedical research
Assemble relevant data
Decision to study
Analyze data
Interpret results
Deliver tailored message
Take action to change practice
Data and information
PERFORMANCE TO DATA
Research during normal care
Analyzed
Problematic factors identified
Action identified
DATA TO KNOWLEDGE
Generate and test hypotheses
Define the problem
Developing algorithms
Learning as (intermittent information exchange between the clinical domain and the research domain / Research influences practice, practice influences research
Learning as a (technology) aided continuous circular process of converting (routine care) data into knowledge, knowledge in to performance, performance to data
Point of care
Evaluate performance of tools and their effect on patient care and outcomes
Implementing tools
Integration in knowledge networks
Validation
Continuous testing of interventions in real-world settings
KNOWLEDGE TO PERFORMANCE
Application of best available evidence (considering patient needs, local system, community context)
Feedback to clinicians, employees, leadership
Actions assigned, performed, monitored
Evaluate the effects of implementation to redefine or de-implement ineffective or harmful interventions
Observe own result and compare to others'
Quality improvement through peers changing each others' behavior based on historical performance
Learn from large numbers of patient experiences through electronic health record systems
Learning as a recurrent interaction between stakeholders to reveal/discuss underlying processes and to identify opportunities for change and share best practices
Common repository to share knowledge
High degree of transparency and ongoing sharing of measures and aggregate and site-level performance
Shared situational awareness
Shared standards
Shared knowledge
Shared tools and resources
Sharing ideas, knowledge, and know-how
Monthly calls and meeting to evaluate the network functioning
Evaluate management structure regularly
Networks evaluate their policies in regular cycles

Figure S8. Intermediary visualization of the 'Learning and evaluation (LHS only)' characteristic. *The visualization shows an overview of learning loops with evaluations (where applicable) and their possible relations and enablers resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

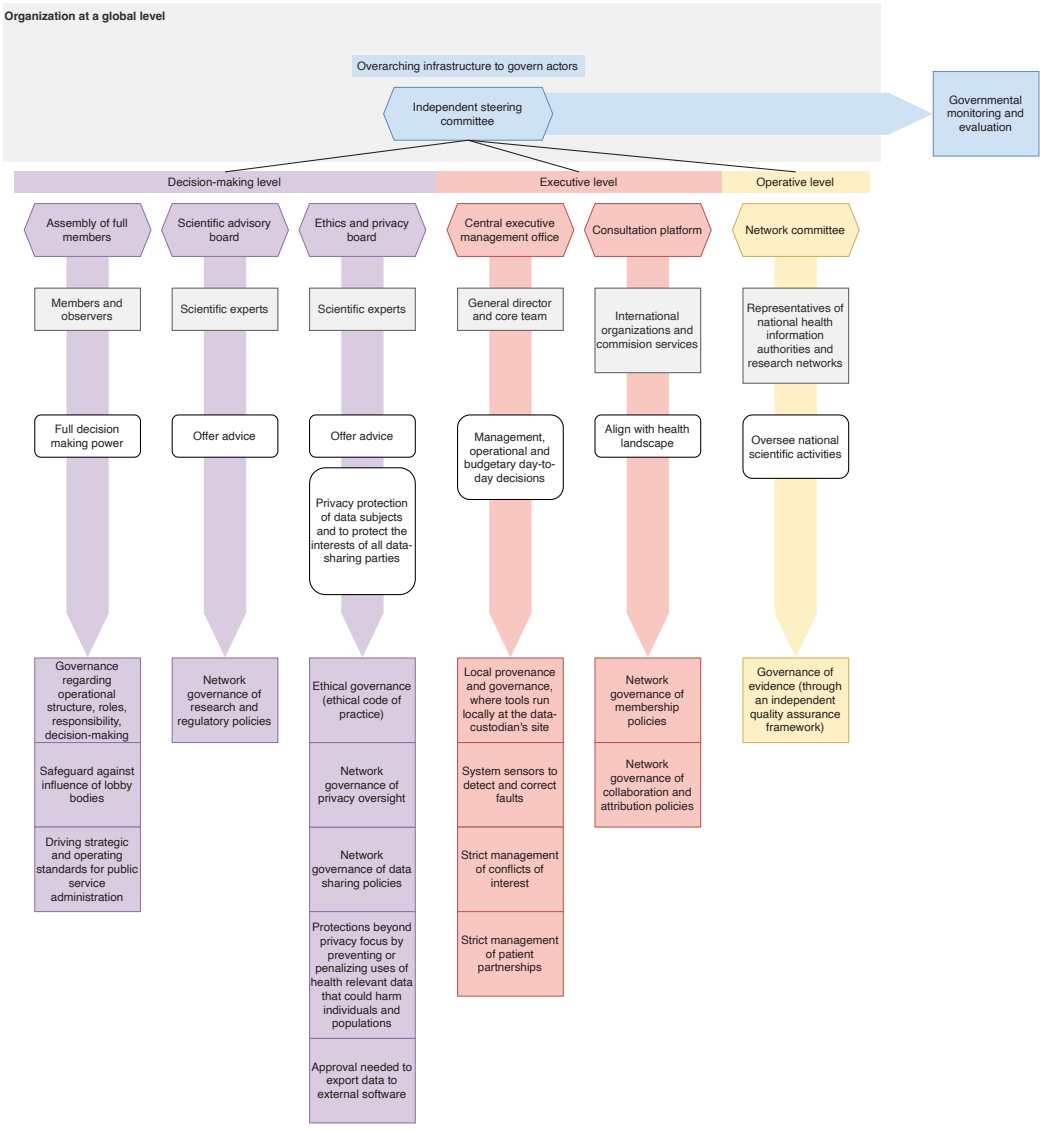


Figure S9. Intermediary visualization of the 'Governance' characteristic. *The visualization shows an overview of a possible governance structure resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

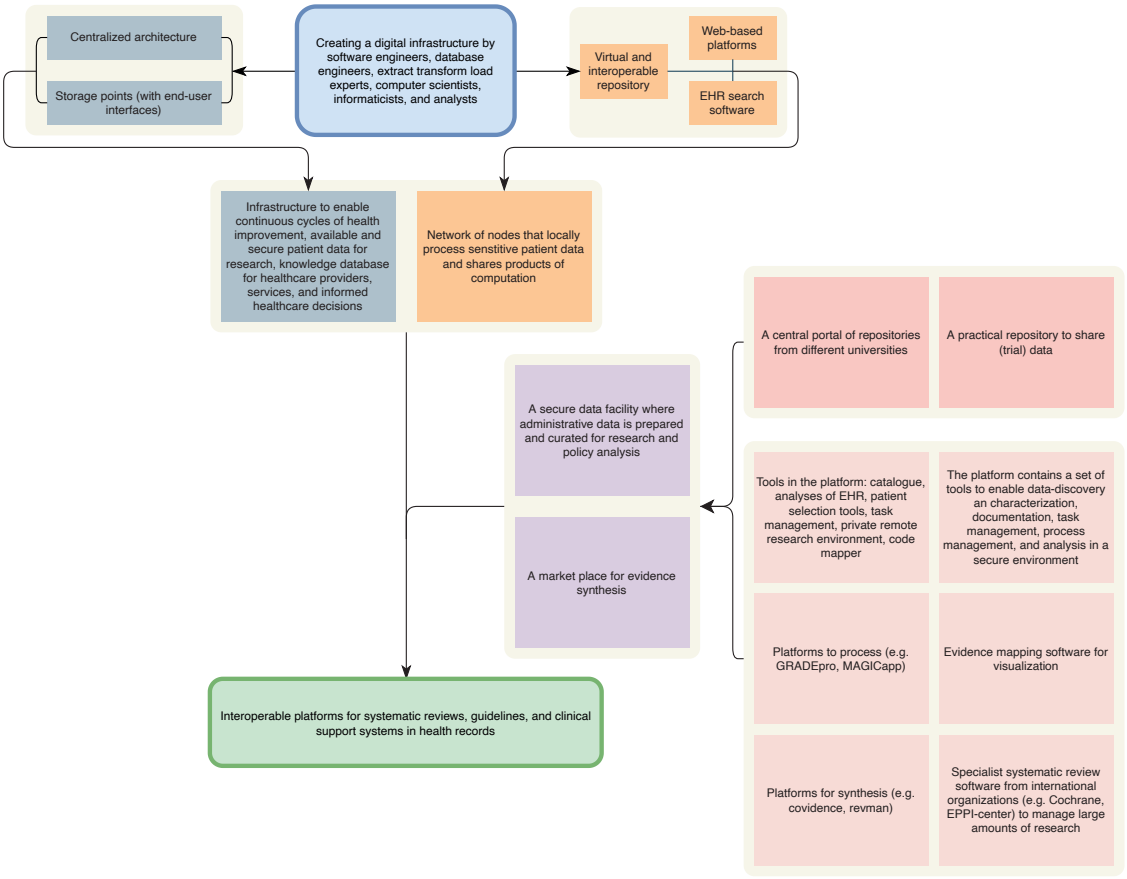


Figure S10. Intermediary visualization of the 'Platforms' characteristic. *The visualization shows an overview of platforms, tools, uses, architecture and their possible relations resulting from the emerging themes and their items identified in the first qualitative intermediary analysis (see Table S3 in Additional File 3).*

ADDITIONAL FILE 4

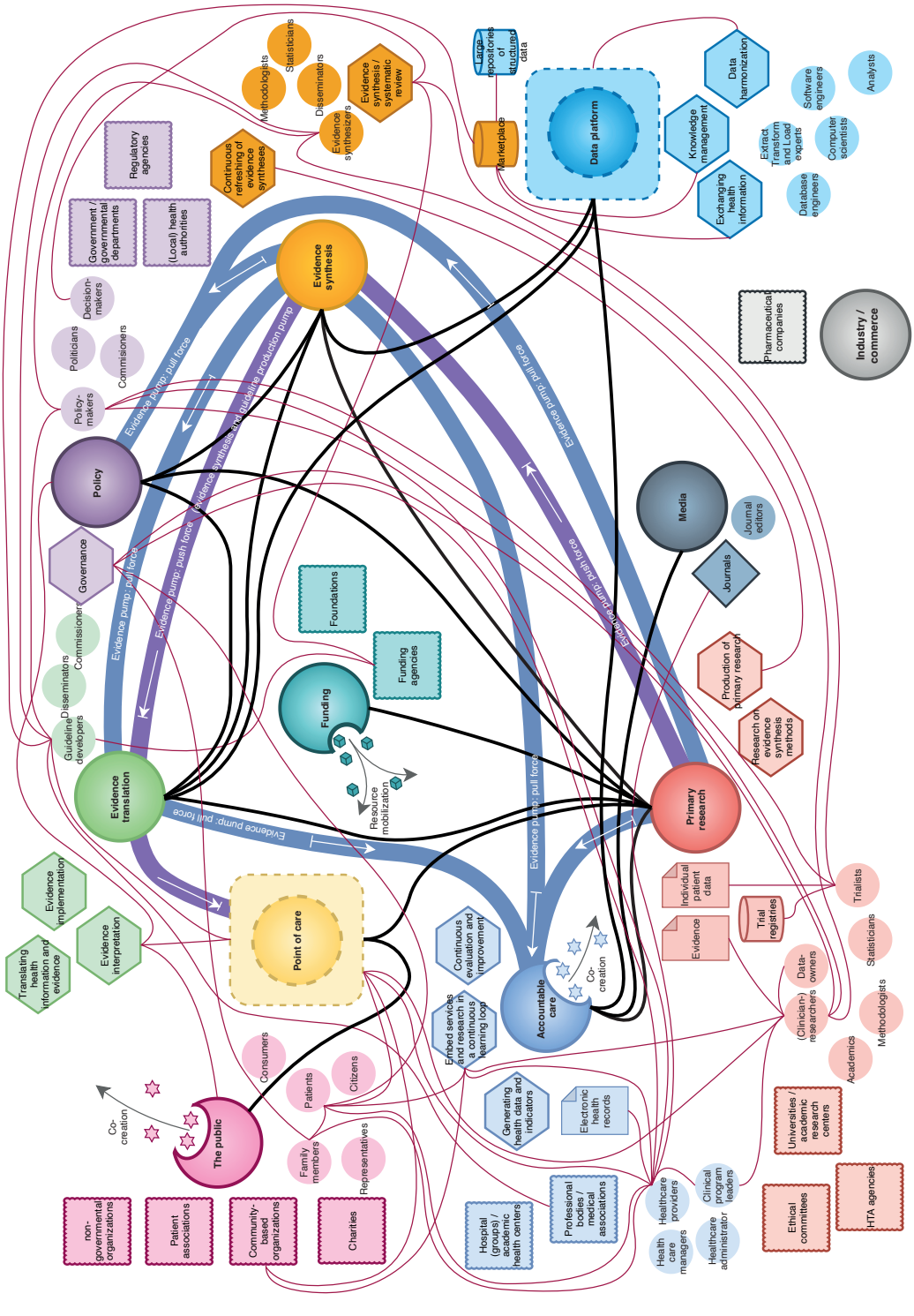


Figure S1. A complex representation of the foundation layer of a symbiotic healthcare ecosystem model. *The visualization shows a complex overview of links, actors, and components as a possible foundation of a symbiotic healthcare ecosystem model. Thin red lines show described links involving small or sub-components in the system, while thick black lines show links between entities. The visualization was based on the extracted data, the preceding intermediary analyses, and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3). Actors (small colored circles), organizations (clouded squares) and components (outlined hexagons) within entities (large colored outlined circles) are added to show a possible foundation of a symbiotic healthcare model. Entities generate data (outlined documents), develop products (outlined diamond), and may store data in databases (outlined cylinders). Co-creation throughout the system is performed by the public and the accountable care entities (small stars), while the funding entity mobilized resources throughout the system (small cubes). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. The flow of evidence is presented as pink [push force] and blue [pull force] ribbons.*

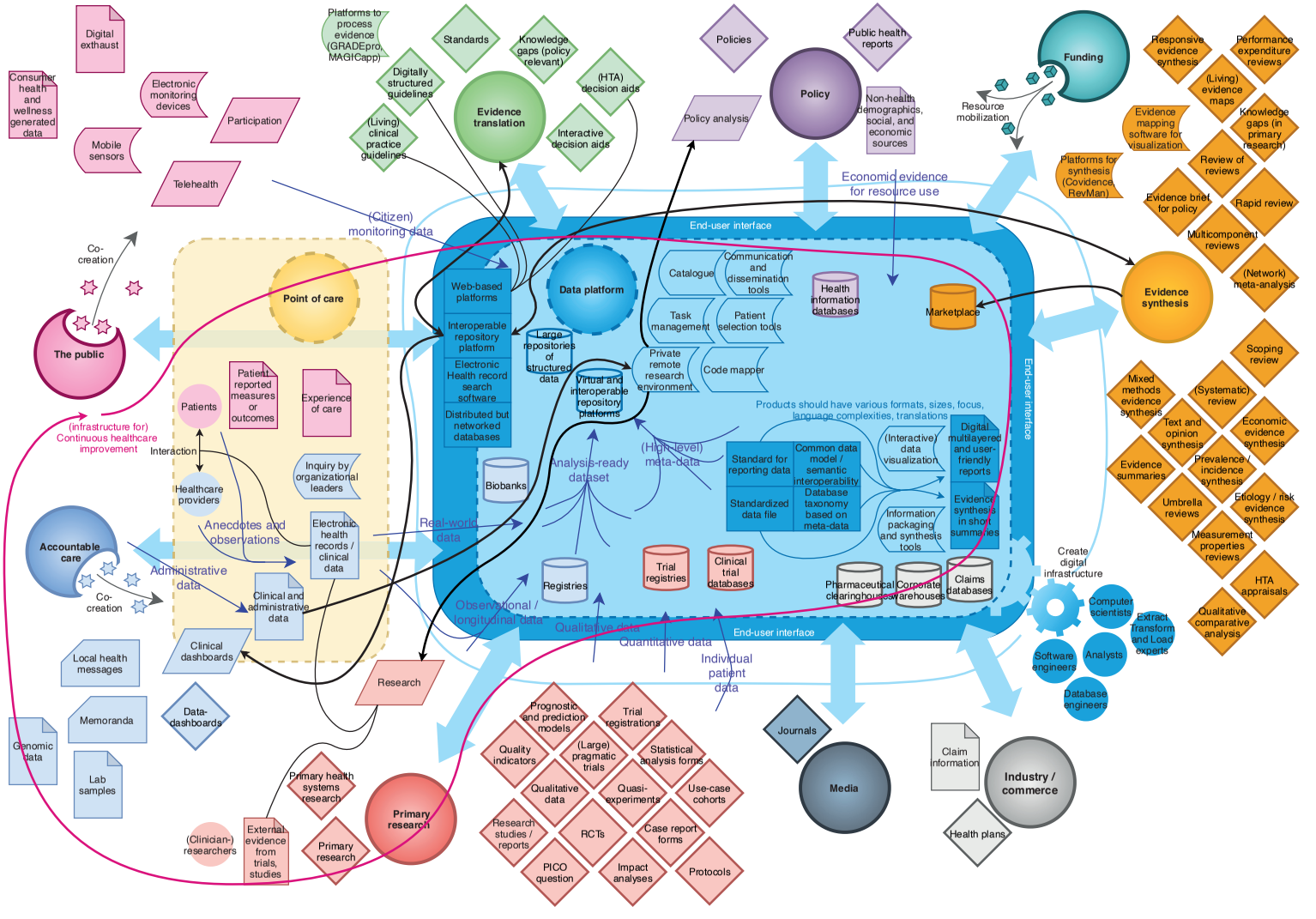

Digital exhaust
Consumer health and wellness generated data
Electronic monitoring devices
Participation
Mobile sensors
Telehealth
Platforms to process evidence (GRADEpro, MAGICapp)
Standards
Knowledge gaps (policy relevant)
Digitally structured guidelines
Evidence translation
(HTA) decision aids
(Living) clinical practice guidelines
Interactive decision aids
Policies
Public health reports
Policy
Policy analysis
Non-health demographics, social, and economic sources
Funding
Resource mobilization
Responsive evidence synthesis
Performance expenditure reviews
(Living) evidence maps
Evidence mapping software for visualization
Knowledge gaps (in primary research)
Platforms for synthesis (Covidence, RevMan)
Review of reviews
Evidence brief for policy
Rapid review
Multicomponent reviews
(Network) meta-analysis
Economic evidence for resource use
(Citizen) monitoring data
Co-creation
The public
Point of care
Patients
Patient-reported measures or outcomes
Experience of care
(infrastructure for) Continuous healthcare improvement
Interaction
Healthcare providers
Inquiry by organizational leaders
Anecdotes and observations
Electronic health records / clinical data
Accountable care
Administrative data
Co-creation
Clinical and administrative data
Clinical dashboards
Local health messages
Memoranda
Data-dashboards
Genomic data
Lab samples
Research
Primary health systems research
Primary research
Primary research
(Clinician-) researchers
External evidence from trials, studies
End-user interface
Web-based platforms
Interoperable repository platform
Electronic Health record search software
Distributed but networked databases
Data platform
Large repositories of structured data
Virtual and interoperable repository platforms
Catalogue
Communication and dissemination tools
Task management
Patient selection tools
Private remote research environment
Code mapper
Health information databases
Marketplace
Products should have various formats, sizes, focus, language complexities, translations
(High-level) meta-data
Analysis-ready dataset
Biobanks
Standard for reporting data
Common data model / semantic interoperability
Standardized data file
Database taxonomy based on meta-data
(Interactive) data visualization
Digital multilayered and user-friendly reports
Information packaging and synthesis tools
Evidence synthesis in short summaries
Real-world data
Registries
Trial registries
Clinical trial databases
Pharmaceutical clearinghouses
Corporate warehouses
Claims databases
Observational / longitudinal data
Qualitative data
Quantitative data
Individual patient data
Evidence synthesis
Scoping review
Mixed methods evidence synthesis
(Systematic) review
Text and opinion synthesis
Economic evidence synthesis
Evidence summaries
Prevalence / incidence synthesis
Umbrella reviews
Etiology / risk evidence synthesis
Measurement properties reviews
HTA appraisals
Qualitative comparative analysis
Create digital infrastructure
Computer scientists
Extract, Transform and Load experts
Software engineers
Analysts
Database engineers
Prognostic and prediction models
Trial registrations
Quality indicators
(Large) pragmatic trials
Statistical analysis forms
Qualitative data
Quasi-experiments
Use-case cohorts
Research studies / reports
RCTs
Case report forms
PICO question
Impact analyses
Protocols
Journals
Media
Claim information
Industry / commerce
Health plans

Figure S2. A complex representation of the infrastructure layer of a SHE model. *The visualization shows a complex overview of data generation and product development by entities and a possible infrastructure for communication, cooperation and continuous improvement using the point of care and the data platform in a symbiotic healthcare (SHE) model. The visualization was based on the extracted data, the preceding intermediary analyses, and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. Entities (large outlined circles) are displayed with their data (documents), products (diamonds), large actions (parallelograms), some actors (small circles), and tools (curved rectangles). Entities connect with the data platform (thick light blue arrows) through a user interface. Several types of data flow to the data platform (purple arrows), where data is stored in databases (outlined cylinders). The data platform has several features and characteristics (outlined deep blue rectangles). The public and accountable care entities will co-create throughout the system (small pink and blue stars) to ensure relevance of processes and products. The funding entity mobilizes resources for processes and products throughout the system (small turquoise cubes). Black lines and arrows show possible links and interactions within the system regarding the model's (data) infrastructure. The magenta arrow shows the area wherein (the infrastructure for) continuous healthcare improvements lie.*

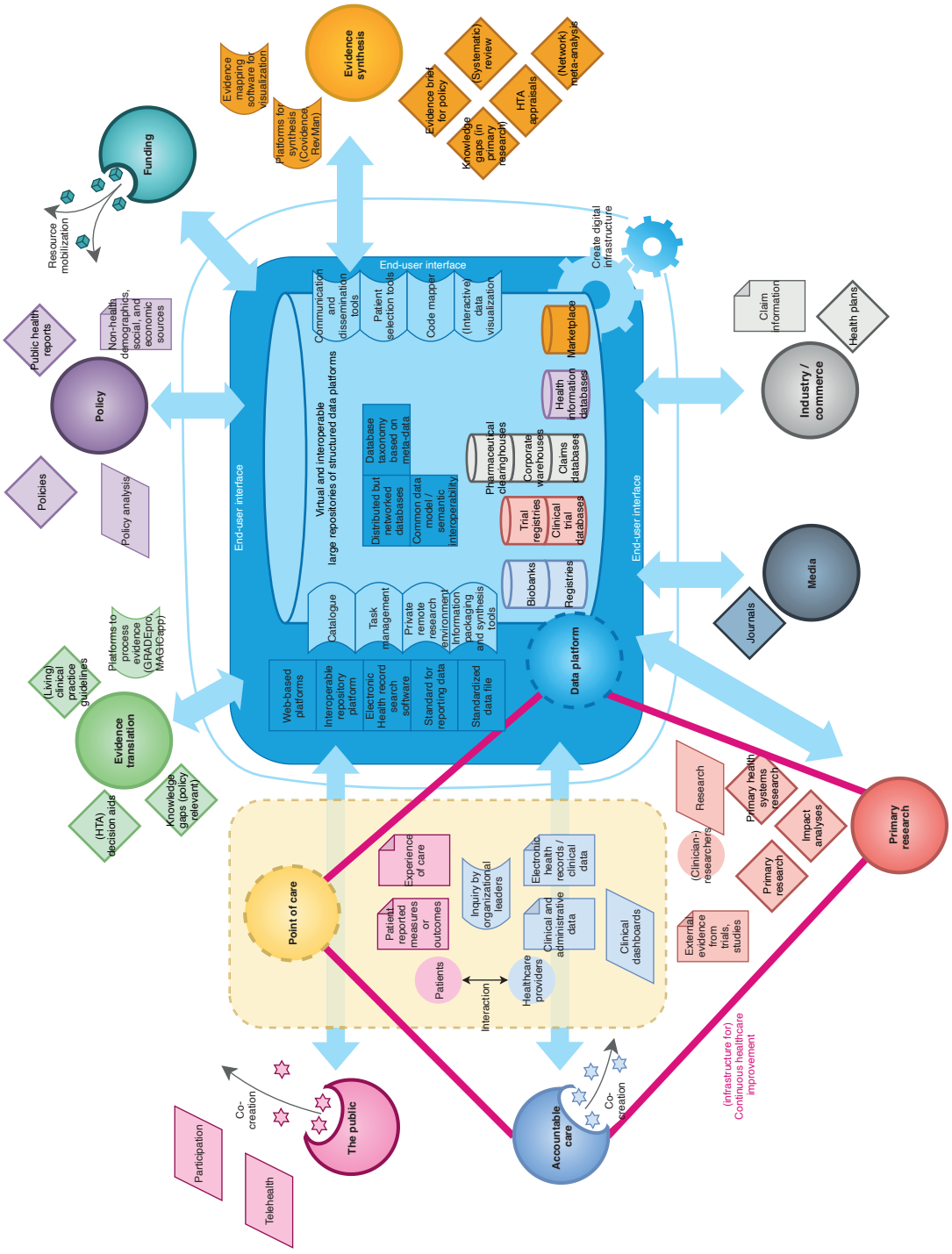
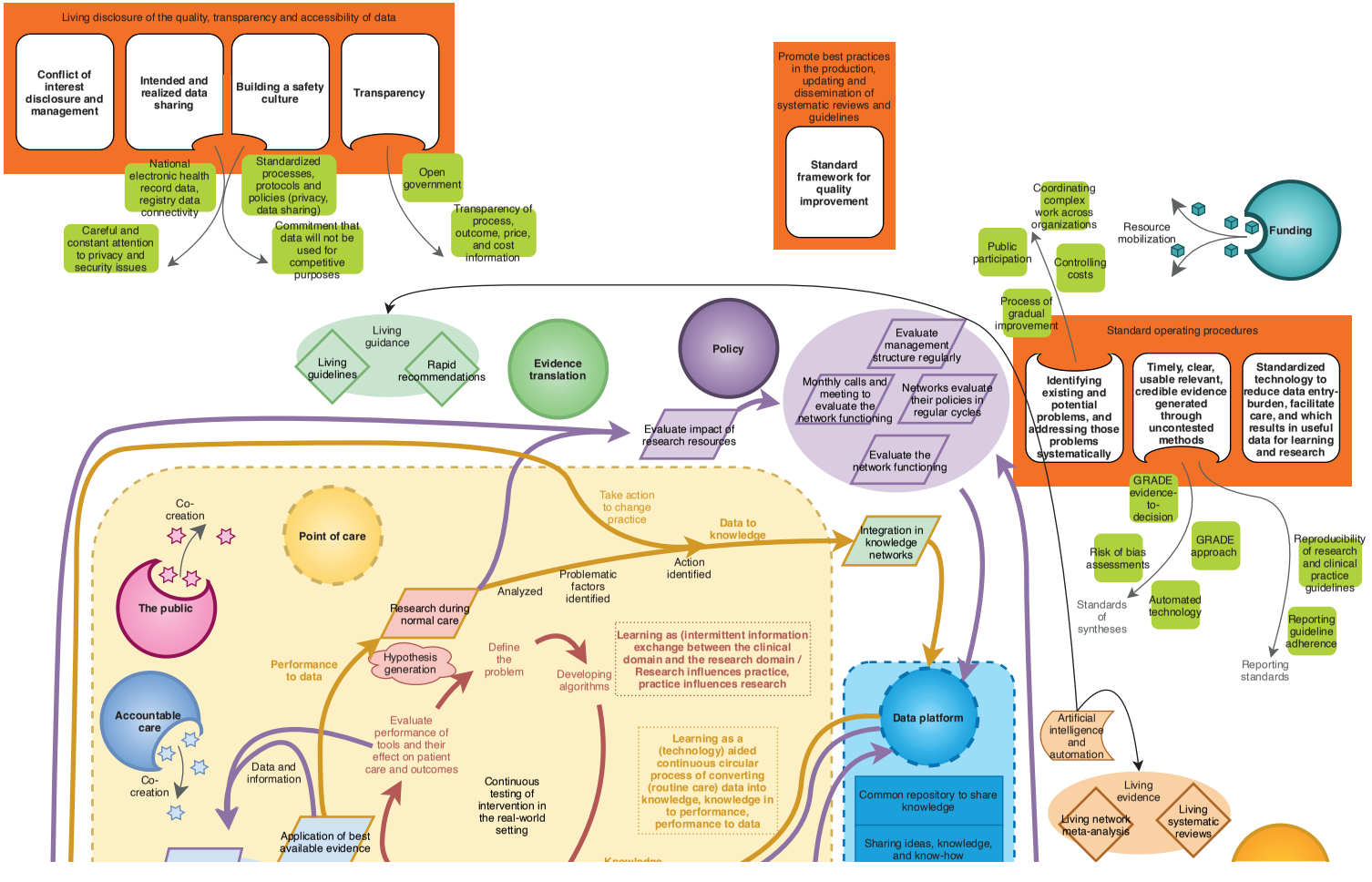

Funding
Resource mobilization
Evidence mapping software for visualization
Platforms for synthesis (Covidence, RevMan)
Evidence synthesis
Evidence brief for policy
(Systematic) review
Knowledge gaps (in primary research)
HTA appraisals
(Network) meta-analysis
Create digital infrastructure
End-user interface
Communication and dissemination tools
Patient selection tools
Code mapper
(Interactive) data visualization
Marketplace
Virtual and interoperable large repositories of structured data platforms
Database taxonomy based on meta-data
Distributed but networked databases
Common data model / semantic interoperability
Pharmaceutical clearinghouses
Corporate warehouses
Claims databases
Health information databases
Trial registries
Clinical trial databases
Biobanks
Registries
Catalogue
Task management
Private remote research environment
Information packaging and synthesis tools
Web-based platforms
Interoperable repository platform
Electronic Health record search software
Standard for reporting data
Standardized data file
Data platform
Public health reports
Non-health demographics, social, and economic sources
Policy
Policies
Policy analysis
Platforms to process evidence (GRADEpro, MAGICapp)
Living clinical practice guidelines
Evidence translation
(HTA) decision aids
Knowledge gaps (policy relevant)
Claim information
Health plans
Industry / commerce
Journals
Media
Point of care
Experience of care
Patient-reported measures or outcomes
Inquiry by organizational leaders
Electronic health records / clinical data
Clinical and administrative data
Clinical dashboards
Patients
Interaction
Healthcare providers
Research
(Clinician-) researchers
Primary health systems research
Impact analyses
Primary research
External evidence from trials, studies
Primary research
The public
Participation
Co-creation
Telehealth
Accountable care
Co-creation
(Infrastructure for) Continuous healthcare improvement

Figure S3. A simplified representation of the infrastructure layer of a SHE model. *The visualization shows a simplified overview of data, communication, and improvement infrastructure in a symbiotic healthcare (SHE) model to increase the comprehensibility of the figure. The visualization was based on the extracted data, the preceding intermediary analyses and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. Entities (large outlined circles) are displayed with their data (documents) products (diamonds), large actions (parallelograms), some actors (small circles), and tools (curved rectangles). Entities connect with the data platform (thick light blue arrows) through a user interface, where data is stored in databases (outlined cylinders). The data platform has several features and characteristics (outlined deep blue rectangles). The public and accountable care entities will co-create throughout the system (small pink and blue stars) to ensure relevance of processes and products. The funding entity mobilizes resources for processes and products throughout the system (small turquoise cubes). The bright magenta lines shows the area wherein (the infrastructure for) continuous healthcare improvements lie.*


Living disclosure of the quality, transparency and accessibility of data
Conflict of interest disclosure and management
Intended and realized data sharing
Building a safety culture
Transparency
National electronic health record data, registry data connectivity
Standardized processes, protocols and policies (privacy, data sharing)
Open government
Careful and constant attention to privacy and security issues
Commitment that data will not be used for competitive purposes
Transparency of process, outcome, price, and cost information
Promote best practices in the production, updating and dissemination of systematic reviews and guidelines
Standard framework for quality improvement
Coordinating complex work across organizations
Public participation
Controlling costs
Resource mobilization
Funding
Process of gradual improvement
Standard operating procedures
Identifying existing and potential problems, and addressing those problems systematically
Timely, clear, usable relevant, credible evidence generated through uncontested methods
Standardized technology to reduce data entry-burden, facilitate care, and which results in useful data for learning and research
Living guidance
Living guidelines
Rapid recommendations
Evidence translation
Policy
Evaluate impact of research resources
Evaluate management structure regularly
Monthly calls and meeting to evaluate the network functioning
Networks evaluate their policies in regular cycles
Evaluate the network functioning
GRADE evidence-to-decision
GRADE approach
Risk of bias assessments
Automated technology
Standards of syntheses
Reproducibility of research and clinical practice guidelines
Reporting guideline adherence
Reporting standards
Co-creation
Point of care
The public
Take action to change practice
Data to knowledge
Integration in knowledge networks
Action identified
Problematic factors identified
Analyzed
Research during normal care
Performance to data
Hypothesis generation
Define the problem
Developing algorithms
Learning as (intermittent information exchange between the clinical domain and the research domain / Research influences practice, practice influences research
Accountable care
Co-creation
Data and information
Evaluate performance of tools and their effect on patient care and outcomes
Continuous testing of intervention in the real-world setting
Learning as a (technology) aided continuous circular process of converting (routine care) data into knowledge, knowledge in to performance, performance to data
Application of best available evidence
Data platform
Common repository to share knowledge
Sharing ideas, knowledge, and know-how
Artificial intelligence and automation
Living evidence
Living network meta-analysis
Living systematic reviews

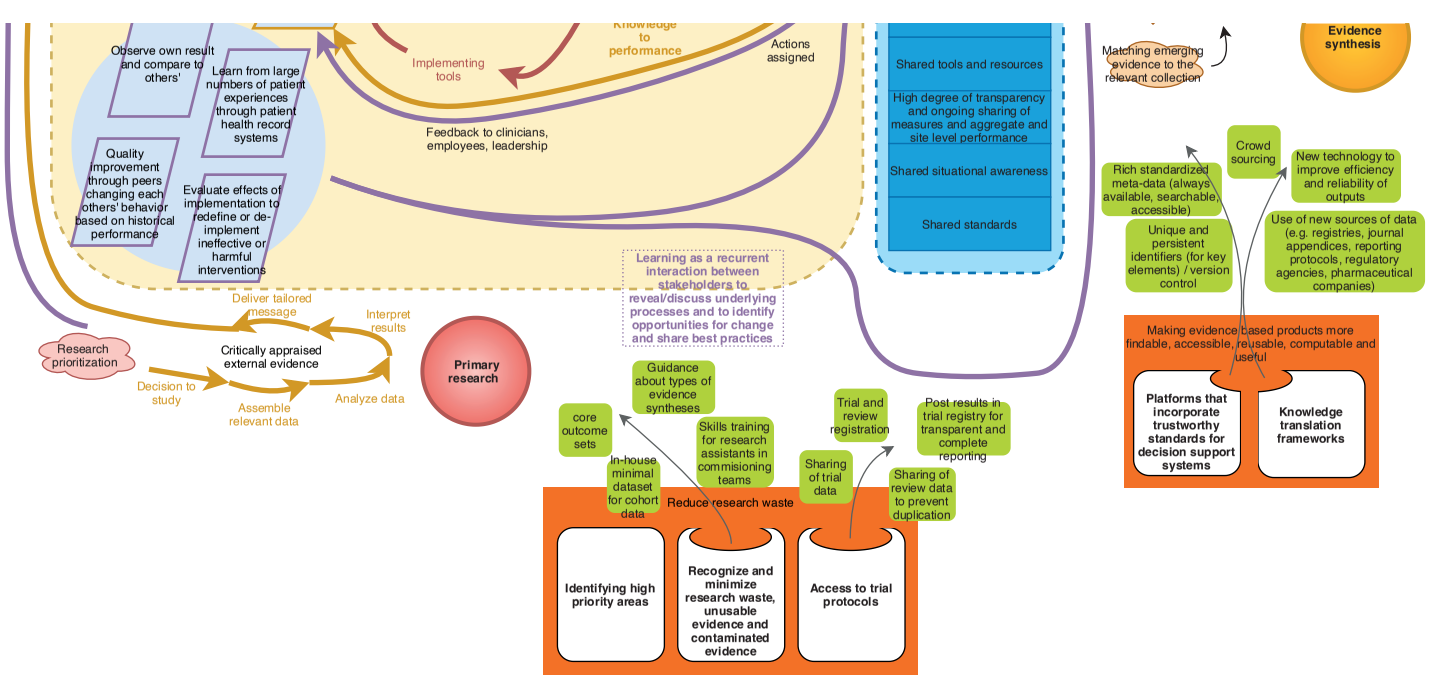


Figure S4. A complex representation of the quality, learning, and evaluation layer of a SHE model. *The visualization shows a complex overview of possible processes and values for learning and quality in a symbiotic healthcare (SHE) model. The visualization was based on the extracted data, the preceding intermediary analyses and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. The data platform has several features and characteristics (outlined deep blue rectangles). Entities (large outlined circles) are displayed with some products (diamonds), large actions (parallelograms), small actions (clouds), some actors (small circles), and tools (curved rectangles). The public and accountable care entities will co-create throughout the system (small pink and blue stars) to ensure relevance of processes and products. Three evaluation and learning processes are displayed as a loop of deep yellow arrows (data to knowledge, knowledge to performance, performance to data), a bilateral red arrow (interaction between research and practice), and a loop of light purple arrows (interactions between stakeholders). The deep orange fields are values from which quality actions (green fields) arise throughout the system to ensure trustworthy processes and products.*

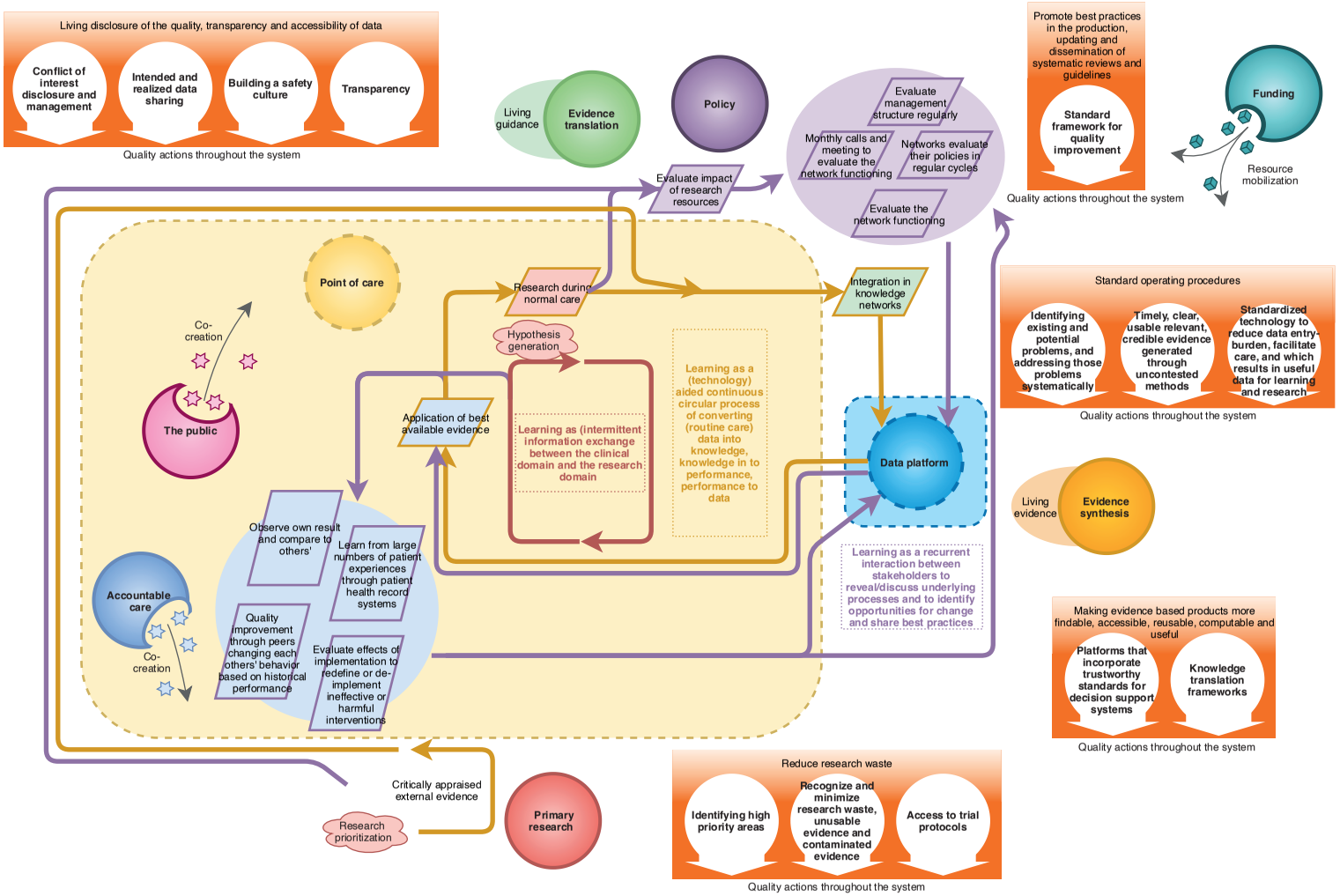
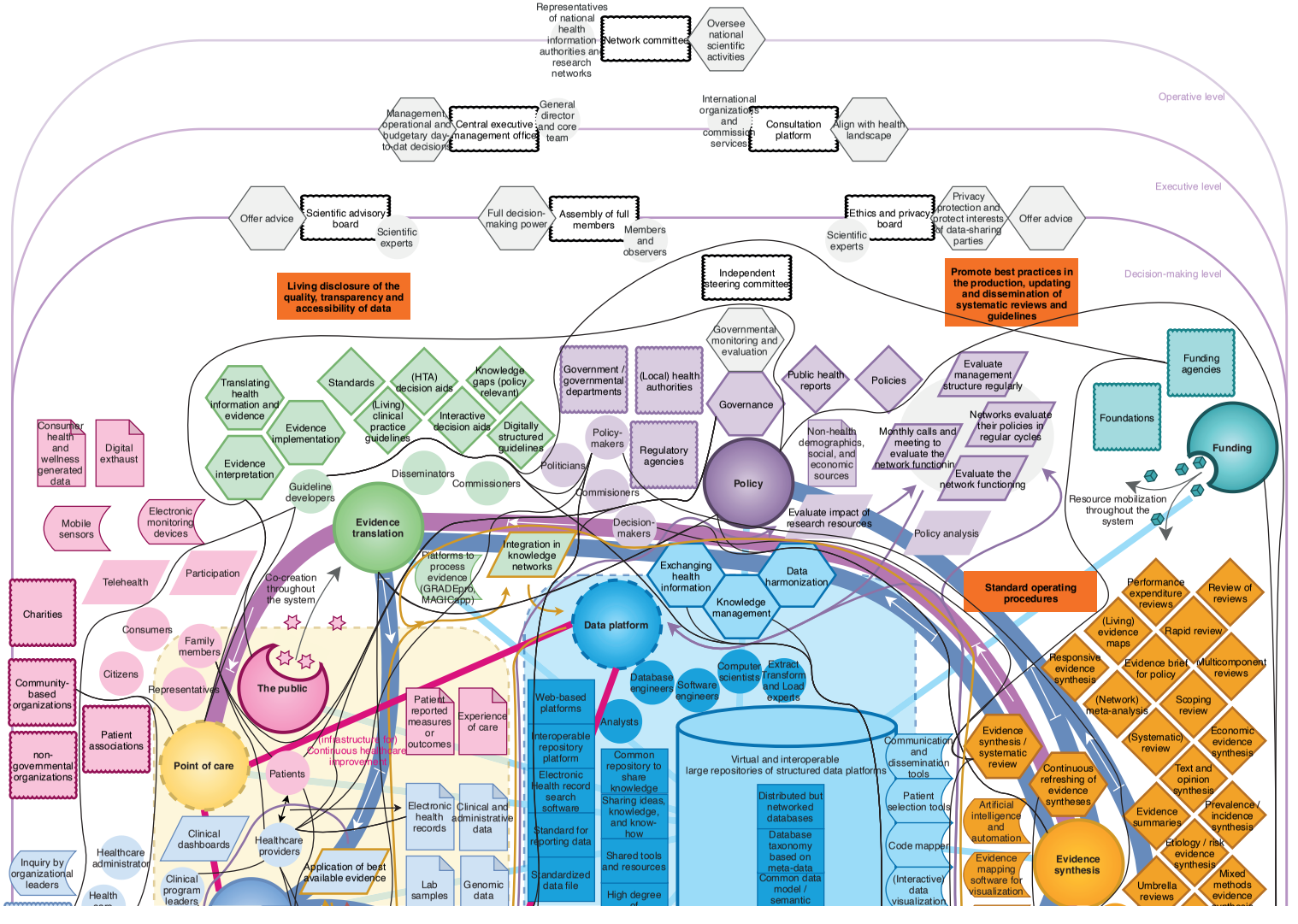

Living disclosure of the quality, transparency and accessibility of data
Conflict of interest disclosure and management
Intended and realized data sharing
Building a safety culture
Transparency
Quality actions throughout the system
Living guidance
Evidence translation
Policy
Evaluate impact of research resources
Evaluate management structure regularly
Monthly calls and meeting to evaluate the network functioning
Networks evaluate their policies in regular cycles
Evaluate the network functioning
Promote best practices in the production, updating and dissemination of systematic reviews and guidelines
Standard framework for quality improvement
Quality actions throughout the system
Funding
Resource mobilization
Point of care
Research during normal care
Integration in knowledge networks
Co-creation
The public
Hypothesis generation
Application of best available evidence
Learning as (intermittent information exchange between the clinical domain and the research domain
Learning as a (technology) aided continuous circular process of converting (routine care) data into knowledge, knowledge in to performance, performance to data
Data platform
Standard operating procedures
Identifying existing and potential problems, and addressing those problems systematically
Timely, clear, usable relevant, credible evidence generated through uncontested methods
Standardized technology to reduce data entry-burden, facilitate care, and which results in useful data for learning and research
Quality actions throughout the system
Observe own result and compare to others'
Learn from large numbers of patient experiences through patient health record systems
Accountable care
Quality improvement through peers changing each others' behavior based on historical performance
Evaluate effects of implementation to redefine or de-implement ineffective or harmful interventions
Co-creation
Learning as a recurrent interaction between stakeholders to reveal/discuss underlying processes and to identify opportunities for change and share best practices
Living evidence
Evidence synthesis
Making evidence based products more findable, accessible, reusable, computable and useful
Platforms that incorporate trustworthy standards for decision support systems
Knowledge translation frameworks
Quality actions throughout the system
Critically appraised external evidence
Research prioritization
Primary research
Reduce research waste
Identifying high priority areas
Recognize and minimize research waste, unusable evidence and contaminated evidence
Access to trial protocols
Quality actions throughout the system

Figure S5. A simplified representation of the quality, learning, and evaluation layer of a SHE model. *The visualization shows a simplified overview of possible processes and values for learning and quality in a symbiotic healthcare (SHE) model to increase the comprehensibility of the figure. The visualization was based on the extracted data, the preceding intermediary analyses and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. Entities (large outlined circles) are displayed with large actions (parallelograms) and small actions (clouds). The public and accountable care entities will co-create throughout the system (small pink and blue stars) to ensure relevance of processes and products. Three evaluation and learning processes are displayed as a loop of deep yellow arrows (data to knowledge, knowledge to performance, performance to data), a bilateral red arrow (interaction between research and practice), and a loop of light purple arrows (interactions between stakeholders). The deep orange fields are values throughout the system to ensure trustworthy processes and products.*

Governance regarding operational structure, roles, responsibility, decision-making
Safeguard against influence of lobby bodies
Driving strategic and operating standards for public service administration
Local provenance and governance, where tools run locally at the data-custodian's site
System sensors to detect and correct faults
Strict management of conflicts of interest
Strict management of patient partnerships
Policy
Governmental monitoring and evaluation
Network governance of research and regulatory policies
Central executive management office
Management, operational and budgetary day to-dat decisions
General director and core team
Assembly of full members
Full decision-making power
Members and observers
Scientific advisory board
Offer advice
Scientific experts
International organizations and commission services
Consultation platform
Align with health landscape
Ethics and privacy board
Offer advice
Privacy protection and protect interests of data-sharing parties
Scientific experts
Network governance of membership policies
Network governance of collaboration and attribution policies
Ethical governance (ethical code of practice)
Network governance of privacy oversight
Network governance of data sharing policies
Protections beyond privacy focus by preventing or penalizing uses of health relevant data that could harm individuals and populations
Approval needed to export data to external software
Organization at a global level
Independent steering committee
Overarching infrstructure to govern actors
Operative level
Executive level
Decision-making level
Representatives of national health information authorities and research networks
Network committee
Oversee national scientific activities
Governance of evidence (through an independent quality assurance framework)

Figure S6. A visual representation of the governance layer of a SHE model. *The visualization shows an overview of a possible governance structure in a symbiotic healthcare (SHE) model. The visualization was based on the extracted data, the preceding intermediary analyses and the gained understanding from these analyses (see Tables S1-S3 and Figures S1-S10 in Additional File 3).The policy entity's (large outlined circle) possible policy structure is displayed on several levels: the global level, the decision-making level, the executive level, and the operative level. Boards and committees (clouded rectangles) with their actors (small circles) and components (hexagons) are displayed. Their possible activities are displayed in outlined boxes.*

ADDITIONAL FILE 5



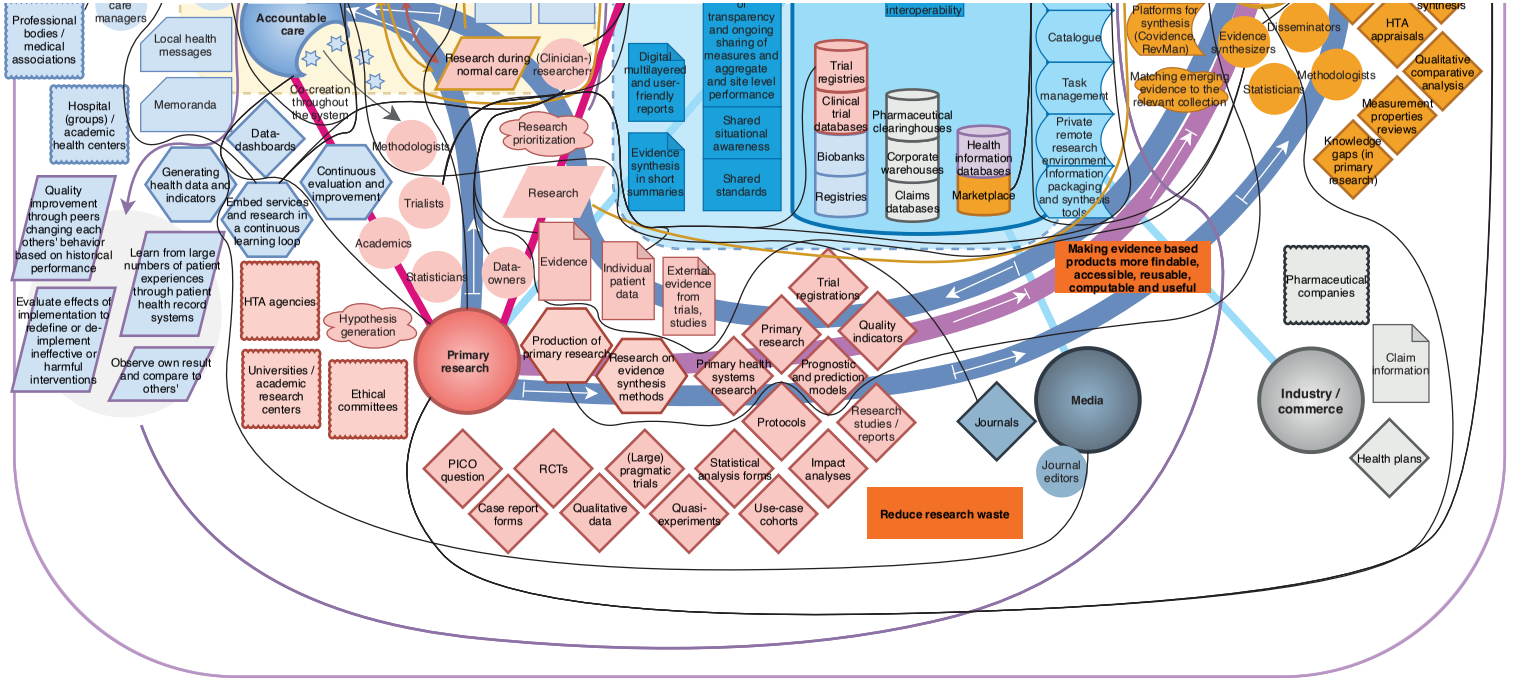


Figure S7. A complex representation of an aggregated SHE model. *The visualization shows a complex overview of a possible model for a symbiotic healthcare system (SHE) by combining the four separate layers (See Figures S1-S8 in Additional File). All the entities (large outlined circles) are displayed with their components (outlined hexagons), organizations (clouded squares) and actors (small circles). Several entities generate data (documents), develop products (diamonds), have large actions (parallelograms) and small actions (clouds), and use tools (outlined curved rectangles). Data and information are stored in databases (outlined cylinders). The flow of evidence is presented as pink [push force] and blue [pull force] ribbons. Co-creation throughout the system is performed by the public and the accountable care entities (small stars), while the funding entity mobilized resources throughout the system (small cubes). The point of care (yellow dashed circle) and the data platform (blue dashed circle) are displayed as locations in the system. Entities have access to the data platform (blue lines) which has several features and characteristics (outlined blue rectangles). The infrastructure for continuous healthcare improvement is depicted as a network of thick magenta lines between the accountable care and primary research entities, together with the two locations in the system. Values for quality, evaluation, and learning (bright orange rectangles) will cause entities to perform quality actions (not shown) throughout the system. Finally, three levels of governance are shown (purple outlines of the model), with their possible boards and committees. Three pathways for evaluation and learning processes are displayed as red, yellow, and purple arrows. Black lines and arrows show possible important links and interactions within the system*

CHAPTER 3

A whole system approach for evaluating best practices in healthcare

C.M. Lentz
M.S. Oerbekke
R. Spijker
L. Hooft
M.J. van der Laan

Submitted

ABSTRACT

BACKGROUND

Best practices play a vital role in generating evidence for assessing healthcare outcomes, resilience and wellbeing of health and care workers, and patient experiences. Evaluating best practices is crucial for understanding their processes, characteristics, and influencing factors. Such evaluation is an essential step toward establishing a learning healthcare, where the assimilation of best practices is facilitated by aligning scientific research, informatics, incentives, and cultural norms. This synergy promotes continuous improvement and innovation in healthcare, ultimately leading to better outcomes and patient experiences and resilient care workers. This study aims to propose a framework that uses qualitative methods to identify quality indicators for evaluating best practices in hospitals, facilitating the transfer of successful practices.

METHODS

A hermeneutic literature review process was used to select relevant literature for the creation of a framework to evaluate best practices in the hospital setting. The framework was constructed based on existing frameworks used for analyses related to best practices in healthcare. Several rounds of literature searches were conducted to identify quality indicators. The authors used a qualitative approach to construct the framework, and inductive thematic analysis was performed to identify emerging themes from the quality indicators identified in the literature. In addition, we interviewed an expert in the field to validate the literature results.

RESULTS

The evaluation of best practices in a hospital environment takes place on four essential axes: context, processes, outcomes and learning capability. The framework shows that within these four axes, a multitude of criteria, themes, sub-themes and proposed quality indicators exist to appraise best practices. Thematic analysis revealed that culture, resources and learning were recurring themes across all four axes.

CONCLUSIONS

The framework for evaluating best practices in hospitals serves as a valuable tool with the potential to enhance the quality of healthcare across various practices. It identifies key factors like culture, resources, and learning abilities that drives the success of best practices. While the framework also proposes quality indicators, additional research is needed for their validation. Overall, the framework sheds light on key elements of best practices and provides a practical approach to improving healthcare quality.

Key words: best practices, improvement methodology, healthcare quality

1. BACKGROUND

A universal goal in healthcare is to create the most optimal and efficient healthcare system. Tools to reach this goal are critical reflection on the organization as well as processes, aiming to establishing best practices. Best practices in a hospital setting, characterized by superior outcomes, play a pivotal role in generating practice-based evidence for healthcare outcomes and patient experiences.[1] These practices offer valuable real-world data for assessing efficacy, especially for patient groups excluded from randomized-controlled trials.[1] Practice-based and research-based evidence are complimentary in a learning health system, with research based-evidence informing policies and practice-based evidence providing real-world data regarding said policies.[1]

Recurrent interactions between stakeholders in a healthcare system can facilitate learning by identifying underlying values and opportunities for change, setting common goals, evaluating processes, and sharing best practices.[2] Sharing what constitutes a best practice can improve the quality of healthcare across different practices. Evaluating best practices is vital within a learning healthcare system, allowing for the transfer of effective processes and characteristics to other contexts. These factors are expected to relate to all aspects of the healthcare trajectory. The resulting transcript of the best practice can capture a wide range of aspects and factors that enable it to perform well, including processes and characteristics.

Learning from best practices is instrumental in creating a learning healthcare system that integrates science, informatics, incentives, and culture for continuous improvement and innovation.[3] In such a learning system, new knowledge is generated as a by-product of patients' and families' care experience.[3] While there is currently no consensus on the definition of a best practice, maximizing favourable outcomes whilst minimizing unfavourable outcomes are essential components.[1]

Understanding how to effectively evaluate best practices in hospitals, including the critical processes, characteristics, and factors to consider, remains a challenge. To enhance healthcare quality across the system, a systematic approach is imperative. This study aims to propose a framework that uses qualitative methods to identify potential quality indicators for evaluating best practices in hospitals, offering insights into their success and aiding their transferability.

2. METHODS

2.1 DEFINITIONS

The following definition of 'best practices' in healthcare is based on the definition by Ng et al. and was used in this review: "*Best practices are practices which have shown evidence of*

superior outcomes. A best practice is context dependent and has been assessed in terms of efficiency, social and ethical values, best available evidence, and relevance. A best practice is context dependent but may also be transferable to other practices. This depends on a clear definition of the context, sustainability, and participation of stakeholders."[1]

2.2 THE HERMENEUTIC REVIEW PROCESS

A hermeneutic literature review approach was utilized in the present study to select and analyze pertinent literature. This method is iterative and dynamic, involving multiple rounds of literature search and analysis until sufficient knowledge saturation is attained to address the research question.[4] Unlike a systematic review, the goal of a hermeneutic review is not to comprehensively collect all literature on a particular subject. Rather, the objective is to iteratively acquire adequate information to address the research question. Following the initial literature search and selection, knowledge gaps were identified, and subsequent searches and literature selections were performed to fill these gaps until saturation was achieved. In our current investigation, the author team concurred that knowledge saturation had been achieved once five relevant articles failed to provide any new information in their entirety regarding quality indicators associated with a criterion in the framework.

2.3 CONSTRUCTION OF THE FRAMEWORK

The first step in creating a framework for the evaluation of a best practice in the hospital setting was identifying relevant axes and criteria for the framework. The Medline and Embase databases were first searched (see Additional File 1) from 2018 to April 1st 2021 for frameworks regarding the identification, implementation and/or evaluation of best practices within healthcare. Both evidence and consensus-based frameworks were included in this study. Consensus-based frameworks use methods such as expert panels or the Delphi method to construct their framework, whereas evidence-based frameworks use results from previous studies to construct their framework.

Titles and abstracts of the articles retrieved by the search strategy were screened to select potentially relevant studies by authors MO and CL. The potentially relevant articles were thereafter read full-text by authors MO and CL and selected when containing frameworks regarding the identification, implementation and/or evaluation of best practices within healthcare. Relevant categories and criteria for the evaluation of a best practice within the hospital setting were extracted from the included literature, and subsequently a framework for the evaluation of best practices within the hospital setting was constructed using a qualitative approach. MO and CL independently identified relevant categories and criteria from existing frameworks. A discussion followed to reach consensus on selecting the definitive categories and criteria. Certain similar categories and/or criteria were present in several frameworks found in the literature and were merged for use our framework if relevant.

2.4 IDENTIFICATION OF QUALITY INDICATORS

Once a framework for the evaluation of best practices was constructed, subsequent literature searches were performed to identify quality indicators that could be used to

evaluate processes, characteristics, and aspects of a best practice following the constructed framework.

Several rounds of literature searching were performed in the period of April- September 2021. A literature search for each criterion in the framework was separately performed to identify possible quality indicators pertaining to that criterion. The Medline, Embase, Scopus, Google scholar and Web of Science databases were searched for literature on 'best practices' and 'high performing teams' in combination with search terms regarding factors associated with success and evaluation. Titles and abstracts were screened from the search retrieval to select potentially relevant studies by authors MO and CL. In addition, websites of key healthcare programmes and organisations were searched for existing guides, manuals and other documents concerning evaluation criteria for best practices. The potentially relevant articles were thereafter read full-text by authors MO and CL and selected when: studies contained information regarding at least one quality indicator for the relevant criterion. Studies describing theoretical models and validation studies regarding factors associated with best practices were included in this review. Literature concerning best practices outside the hospital setting were excluded.

2.5 THEMATIC ANALYSIS

Once quality indicators were identified and extracted from the literature, inductive thematic analysis was carried out to identify emerging themes. MO and CL both conducted thematic analysis independently using NVivo version 1.6.2 after which a discussion followed to establish the definitive themes.

3. RESULTS

The literature search for frameworks regarding the identification, implementation and/or evaluation of best practices within healthcare yielded 69 hits. Based on screening of title and abstract, 54 studies were excluded as they did not seem to concern the identification, implementation and/or evaluation of best practices within healthcare. Fifteen articles were selected for full-text evaluation. After this evaluation, 11 more frameworks were excluded, see PRISMA flowchart (Additional File 2). Reasons for exclusion were: not meeting the research question, frameworks focusing solely on medical education, and did not concern a framework. Four studies containing four frameworks reported relevant factors. Additional sources yielded one additional framework. The factors included and excluded from the five frameworks are described in Table 1 and includes whether we considered the framework being evidence or consensus based. One framework (Ng, 2015[1]) was based on evidence found in literature through a systematic literature review. The other frameworks were either a combination or consensus based.

3.1 FRAMEWORK SHOWING QUALITY INDICATORS OF BEST PRACTICES

Based upon the criteria identified in Table 1, a framework for the evaluation of best practices within the hospital setting was constructed. Four axes were identified: context, process, outcomes, and learning and improvement. Context was divided into two criteria based upon

Table 1- Criteria included and excluded from frameworks relevant to best practices

Title	Criteria included	Method of measuring domains	Criteria excluded	Evidence or consensus based?
Framework for selecting best practices in public health[1]	Criteria across context, process, and outcome. Context: - Relevant to needs - Relevant to setting Process: - Community participation (renamed to stakeholder participation) - Stakeholder collaboration (under stakeholder participation) - Use of best available evidence (this criterion is mentioned in the article by Ng et al, but not included in their framework) Outcomes: - Effective (renamed to clinical outcomes, as these are the desirable outcomes within hospital healthcare) - Efficient (renamed to process outcomes) - Sustainability	Context: Evaluate disease burden, conducting needs assessment. Evaluate existing programmes. Process: Stakeholder participation: Appropriate representation of groups. New approaches due to pooling of resources. Outcomes: Clinical: reduce morbidity and mortality, monitoring. Process: Cost-benefit analysis and cost-effectiveness analysis. Sustainability: Training locals, community awareness and self-financing of intervention by community.	Process: - Ethically sound (excluded as this is more relevant to public health, not best practices within hospitals) - Replicable (excluded as replicability depends on context factors, stakeholder participation and sustainability, so covered under these criteria)	Evidence based (systematic literature review)
A Qualitative Study to Develop a Privacy and Nondiscrimination Best Practice Framework for Personalized Wellness Programs[5]	Culture of trust (including data governance and non-discriminative practices)- included within the context-related criteria 'relevant to setting'	/	Culture of health or wellness- excluded as this relates to health or wellness programs for employees which is not directly related to best practices of hospital healthcare	Consensus based (based on consensus of public and legislative stakeholders in the US)
Community pharmacy-based injectable naltrexone service delivery models and best practices[6]	- Leadership support- this criterion is included both within the context-related criteria organizational support and within the process-related criteria team functioning and composition - Establishing referral relationship- covered under the process-related sub-criteria 'interdisciplinary communication'	/	Creating the required infrastructure- this would be relevant for the implementation of a best practice, not for the evaluation of one however Other factors were excluded as they related to implementation of a best practice and were specific to the delivery of naltrexone injections, these include: establishing lab protocols,	Consensus based

			training to administer injections (training of staff is included in the proposed framework in this study, not to this level of specificity however), changing of workflow, activities before, during and after seeing the patient.	
Developing a model of best practice for teams managing crisis in people with dementia: a consensus approach[7]	Service purpose: is included under process-related criteria. Team values: is included under process-related criteria. Reflexivity: is included under learning & improvement Coordination of service: is included under process-related criteria. Decision making: is included under process-related criteria. Outcomes: are included as a major criterion. Responsiveness of the service: is included under context-related criteria. Staffing the service: is included under process-related criteria. Leadership: is included under process-related criteria. Supervision and training: is included under process-related criteria. Joint working and referrals: is included under process-related criteria (interdisciplinary communication) Team base environment: is included under context-related criteria relevant to needs. Assessments: is include under learning & improvement (evaluation of performance). Workload: included under context-related criteria relevant to setting. Stakeholder satisfaction: included under process-related criteria. Quality improvement initiatives: include under learning & improvement	Checklist of activities relating to care of dementia patients.	Accessibility of service: is excluded as this is more relevant for best practices in public health, this study looks at best practices once a patient has already accessed the service (hospital care). Psychosocial and pharmacological interventions: excluded as this is highly specific to care of dementia.	Combined (Systematic review to gather evidence, then consensus-based approach for selection and finally refining standards by field testing and consultation groups)

Best Practice Framework of Fracture Liaison Services in Spain and their coordination with Primary Care[8]	Effective communication: included under process-related criteria (interdisciplinary communication) Clinical reports: included under outcomes (monitoring of performance) Follow up: included under process-related criteria (use of best available evidence)	Brief checklist	/	Consensus based

the framework by Ng et al.:[1] criteria related to needs and criteria related to setting. Process was divided into several sub-criteria based upon findings in literature and criteria mentioned in frameworks described in Table 1. These sub-criteria are: stakeholder participation, team functioning and composition, use of best available evidence and safety. Outcomes were divided into the following sub-criteria: clinical outcomes, process outcomes and sustainability. These are based on those mentioned in the framework by Ng et al.[1] Learning capacity was mentioned in some frameworks (Table 1), terms such as quality improvement initiatives, assessments and reflexivity are used to describe the learning capacity in best practices.

The full framework, including the criteria, (sub)themes, and indicators, are presented in Table 2. Definitions of terms are presented in Additional File 3.

3.2 EMERGING THEMES FROM THE FRAMEWORK

Three themes emerged from the framework across all axes; culture, resources and learning. These are visually presented in Figures 1-3, showing their presence across the axes.

4. DISCUSSION

The aim of this study was to develop a comprehensive framework to evaluate best practices in a hospital setting by conducting a hermeneutic review. The main findings reveal that the evaluation of a best practice involves four essential axes within our framework. Firstly, the context in which a best practice is applied plays a significant role (context axis). Secondly, an evaluation should consider the underlying processes within the best practice (process axis). Next, the outcomes resulting from these processes (outcome axis) and the practice's capacity for learning and adaption must be examined (learning capacity axis). The framework shows that within these four axes, a multitude of criteria, themes, sub-themes and proposed quality indicators to appraise best practices. This underscores the multifaceted nature of the factors contributing to best practices in hospital settings.

Frameworks regarding the identification and dissemination of best practices are reported in current literature.[40] Within public health, evaluation criteria for the selection of a best practice have been described.[1] Public health is, however, a broader concept than hospital healthcare and mainly concerns processes before a patient has reached the hospital. Our framework presents a comprehensive overview of the multitude of factors important in best practices within a hospital setting. Through combining categories and criteria from existing frameworks it was ensured that an extensive overview of evaluation areas was covered in the framework.

Interestingly, recurring themes throughout our framework were culture, resources and learning (Figures 1-3). These themes were present across all four axes, signifying their importance. Culture and learning are often mentioned in learning health systems.[3, 41] As the name suggests, learning is vital in a learning health system, where health care activities are learned from, studied and improved upon.[42] A culture of improvement and innovation is

Table 2 - Framework of quality indicators associated with best practices

Category	Criterion	Themes	Sub-themes	Quality indicators
Context *Contextual factors include characteristics of the target group's circumstances, and its environment. Contextual awareness is needed for interpretation of a best practice. (Based on Ng et al. 2015)[1]* [2]	Relevant to needs *The best practice is relevant to the needs of the target group and stakeholders.* (Based on Ng et al. 2015)[1]	**Organizational support**		Organizational support e.g. in terms of access to expertise, information technology & financial resources[9, 10]
		Stakeholder consensus		Stakeholder consensus to manage best practice recommendation/development and implementation[11-15]
		Needs of the target population		The best practice meets the needs to the target group[12, 13, 16]
	Relevant to setting *Setting describes the environment in which the best practice is implemented. Setting describes environmental factors influencing the ability to reach desired outcomes. (Based on Ng 2015)[1]*	**Culture**		Organizational culture[9, 10] Culture of trust[9]
		Resources		Resources available[17]
		Workload		Workload / (high pressure environment)[18]
Process *A series or network of reproducible steps that together form the basis of action within a best practice*	Stakeholder participation *Participation of partners who enable a sustainable process through complementing perspectives, infrastructure, competencies, and/ or resources. (based on Ng 2015)[1]*	**Decision making**		Engagement of stakeholders in decision making processes[11-15, 19-21]
		Evaluation		Stakeholder involvement in evaluation of best practice[11, 20]
	Team functioning & composition *Team composition entails the members of a team responsible for providing patient care. Team functioning entails the behaviour a team displays.*	**Team culture**	*Non-hierarchical culture*	Collegial interaction rather than hierarchical[22]
			Culture in process	Supporting behaviour[23, 24] Service purpose[24]
		Teamwork	*Communication*	Open communication[22] Closed loop communication[10] Clear/ standardized language[10]

			Interdisciplinary communication[25] Situational awareness[10, 23]
		Leadership	Shared Leadership[23] Leadership support[9] Decision making[25]
		Unity	Shared mental model[10, 24]
		Organizational structure	Dedicated teams[25] Clear roles and responsibilities[21, 25] Coordination of service[25] Interdependence[9]
	Staff competency	*Training & education*	Training & Education of staff[9, 11, 24, 25]
		Skills	Technical skills[10] Non-technical skills[10]
Use of best available evidence *Evidence within medicine is the available body of proof that certain interventions result in desired outcomes, proving efficacy. (Based on Ng 2015)[1]*	**Use of evidence in practice**	*Evidence*	Scientific evidence to support best practice[14, 26] Awareness of evidence-based guidelines[20, 26] Models of care[20, 27]
		Implementation of evidence	Implementation and adherence to best available evidence[26]
		Access to resources	Access to necessary infrastructure and resources[20, 26]
	Culture of improvement		Culture oriented at improvement of care[9, 26]
Safety *Safe clinical practices refer to practices with lowest risk of patient injury.* *(Based on Knox 1999)[22]*	**Safety management**		Managed proactively (leadership)[28] Leader provides updates to team[29] Detection & analysis of injury and near-misses[10] Process design for safety[10]
	Culture of safety		Patient safety is a priority[10, 30]

				Culture of safety[10] Proactive communication (handover)[31]
		Training in process		Emergencies are rehearsed[22, 25]
Outcomes *Outcomes are the measurable effects following an intervention. (Based on Ng 2015)*	Clinical outcomes *Clinical outcomes are the measurable effects on patient health following an intervention. (Based on Ng 2015)[1]*	**Clinical outcomes**		Patient satisfaction[12, 30, 32] Superior clinical outcomes[10, 20, 32]
		Quality measurement		Quality indicators[20, 32]
		Performance measurement		Monitoring of performance (clinical reports)[9]
	Process outcomes (efficiency) *Process outcomes are the measurable effects on efficiency following an intervention. (Based on Ng 2015)[1]*	**Efficiency**	*Human resources efficiency*	Vacancy rate, turnover rate and agency usage[32]
			Efficiency	Performance metrics (e.g., surgical time, turnover time)[33, 34] Use & Incorporation of information technology[9, 34]
			Cost-effectiveness	Cost-effectiveness[20, 32, 34, 35]
		Goal setting		Action plan of short- and long-term goals reflecting team's purpose[9]
	Sustainability *Sustainability describes the long-term continuation of an intervention. (Based on Ng 2015)[1]*	**Sustainability of resources**	*Sustainability of resources*	Availability of financial resources[17]
			Human resources sustainability	Stability of staffing[17, 36, 37]
			Organizational structure in sustainability	Sustainability activities (education, reminders, evaluation)[17, 36]
			Sustainability of training & education	Training of new staff members[17] Educational materials remain available after implementation[11]

			Program development for sustainability	Embedding continuity in program delivery[17, 38, 39]
		Sustainability of culture	*Support base*	Perception of advantages of best practice[36] Incorporating hospital staff in the implementation of a best practice[11]
			Leadership in sustainability	Leadership factors[36]
			Culture in sustainability	Culture of shared accountability[34, 36] Organizational commitment[17, 39]
Learning capacity *Learning capacity describes the ability of a best practice to evaluate its own performance to improve the healthcare provided.*	Learning & Improvement *Learning and improvement entails the ability to learn from past performances and improve through evaluation.*	**Evaluation of performance**		Monitoring and evaluation of performance[11, 20, 34] Debriefing following adverse events[10] Reflexivity[20] Performance indicators to monitor quality of care[14, 20, 34]
		Initiatives		Quality improvement initiatives[10]
		Learning culture		Organizational culture for continuous learning[10]

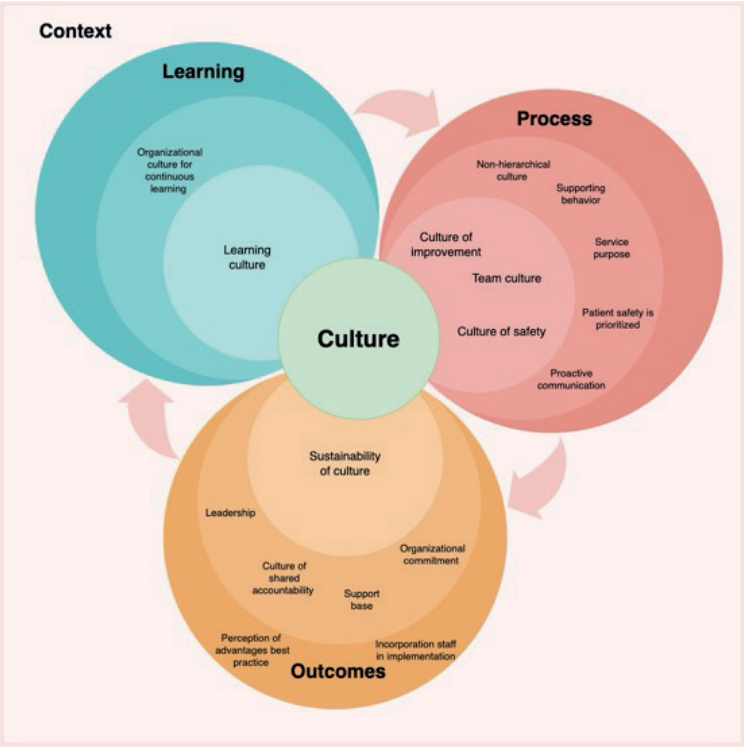


Figure 1. Culture in best practices. *This figure depicts the theme 'culture' as seen in three axes in the framework. Themes and quality indicators related to culture across the three different axes are shown, with the context as background. The arrows indicate how the three axes relate to each other; a process results in certain outcomes which can be used to learn from and adapt the processes.*

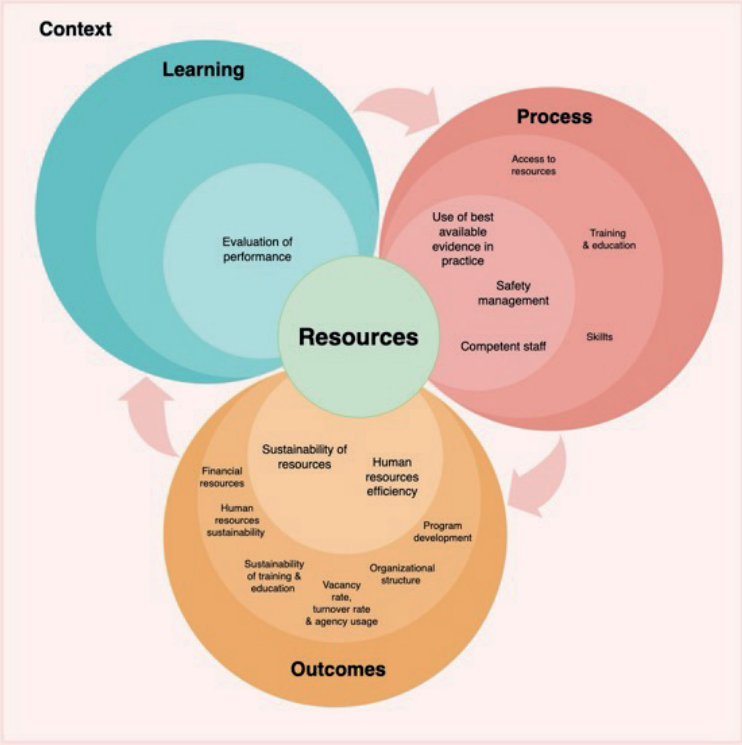


Figure 2. Resources in best practices. *This figure depicts the theme 'resources' as seen in three axes in the framework. This figure shows the widespread importance of resources in best practices, it shows the themes and quality indicators across the three different axes, with the context as background. The arrows indicate how the three axes relate to each other; a process results in certain outcomes which can be used to learn from and adapt the processes.*

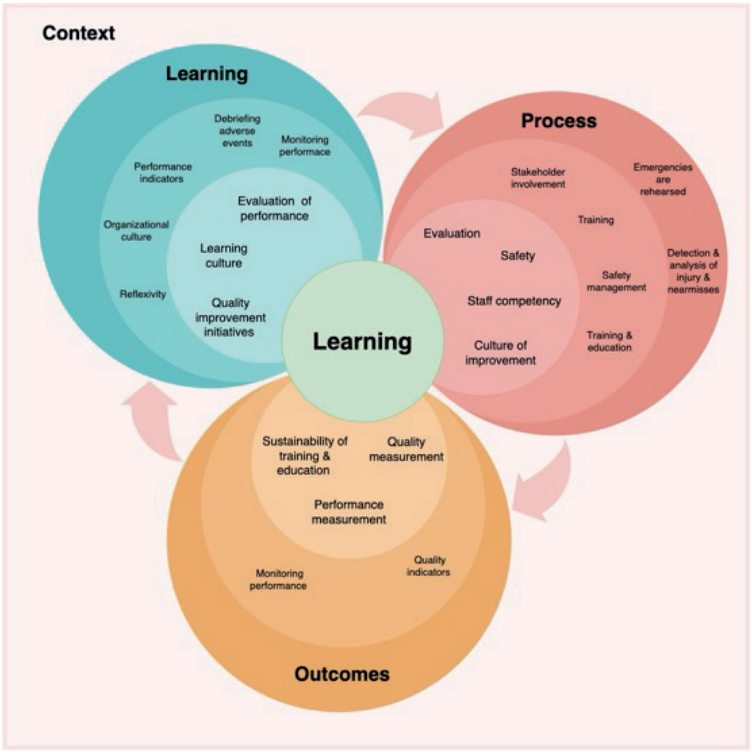


Figure 3. Learning in best practices. *This figure depicts the theme 'learning' as seen in three categories in the framework. This figure shows the widespread importance of learning in best practices, it shows the themes and quality indicators across the three different categories, with the context as background. The arrows indicate how the three axes relate to each other; a process results in certain outcomes which can be used to learn from and adapt the processes.*

reported as an integral part of such a system.[3]

Previous studies have demonstrated the association between organizational and workplace culture and patient outcomes with positive culture linked to lower mortality rates, reduced complications, and increased patient satisfaction.[43] Nevertheless, our framework highlights the pervasive role of culture in a best practice across several axes, as shown in Figure 2. Although previous studies have separately discussed the influence of these cultures on healthcare practices,[44-47] the concept of culture within a best practice along with its multiple facets requires further development and refinement. Notably, 'toxic hospital culture' can severely impact healthcare quality, leading to increased medical errors and staff turnover.[48] Understanding hospital culture is complex, consisting of networks of clinical micro-systems and their associated behaviors and interactions.[48]

Furthermore, the association between healthcare resources and outcomes has been well-established in public health.[49-54] In hospitals in developed countries, strained resources were also found to be associated with elevated mortality, increased complications and extended length of hospital stay.[55] Strained resources refer to situations when there is a discrepancy between supply and demand of various hospital resources, including nurses, doctors, equipment, and beds.[55] These findings illustrate the significance of evaluating the resource axis within the best practice framework, with resources being a recurring theme. Furthermore, resources play a supportive role in culture and learning, particularly through providing outcome data that allowing for learning and improvement.

The importance of learning as a driver for quality improvement is increasingly recognized in healthcare.[3] Learning entails staff training and education, along with the evaluation and learning from performance, as shown in Figure 3. Staff training and education are essential for acquiring technical skills needed for patient care. Moreover, non-technical skills like teamwork and communication have been associated with fewer preventable adverse outcomes.[56-58] Evaluation and learning from performance are essential to achieve quality improvement and is a central concept in both learning health systems and value-based healthcare.[59, 60] Learning health systems are dedicated to ongoing improvement through the use of information, evidence, culture and stakeholder input to implement best practices and generate evidence during the implementation process, creating a cycle of continuous healthcare improvement.[59]

4.1 STRENGTHS AND LIMITATIONS

The strengths of this study include the application of a hermeneutic review methodology, which enabled us to identify a diverse range of topics and indicators to incorporate into the framework, leading to a comprehensive framework for the hospital setting. Furthermore, the framework was constructed through the extraction of distinct domains and criteria from multiple frameworks, resulting in an extensive overview. Thematic analysis of quality indicators allowed for the emergence of themes across the framework, enhancing its coherence and clarity.

Limitations of this study include that saturation was achieved at the criterion level, rather

than at the level of proposed quality indicators. Each proposed quality indicator should be reviewed in greater detail separately to establish the precise data required to fulfill the quality indicator. Additionally, the search strategy for frameworks was restricted to works published from 2018 onwards for reasons of feasibility, which may have resulted in the exclusion of relevant frameworks.

4.2 IMPLICATIONS

Our current study provides a solid base for further research regarding the practical approach of evaluating best practices. We have detailed the categories, criteria, themes and proposed quality indicators that could be included in such an evaluation. Additionally, we have provided a concrete evaluation tool based on our framework, which is available as an additional file (Additional File 4). Additional File 5 shows how the conceptual framework could be used in a real-life evaluation of a best practice. The next steps would include validating the framework across broader and more relevant areas of expertise.

Culture, resources and learning were found to be important themes within the context of best practices. Measuring and evaluating these themes is essential when evaluating a best practice. Resources and learning in hospital settings are quantifiable aspects, but culture, due to its conceptual nature, is however a challenge for measurement. To date, there is no consensus on how to effectively measure culture in healthcare practice.[61] One could theorize that culture might vary across different axes, which could lead to variations in its measurement. However, it remains a subject of debate whether quantifying culture would significantly enhance the evaluation process. Further research is necessary to establish quantifiable quality indicators, allowing for the transition from a conceptual framework to practical application.

5. CONCLUSIONS

In summary, this framework provides a significant step towards improving healthcare quality by shedding light on the key elements of a best practice. It offers a practical approach to evaluate best practices within a hospital setting, ultimately serving as a valuable resource for other clinics and healthcare providers aiming to elevate the quality of care they offer. The framework highlights the multifaceted nature of a best practice, consistently emphasizing key elements such as the culture, resources and learning ability of healthcare organizations. Additionally, it proposes quality indicators associated with best practices, however further research is needed to establish these indicators as quantifiable metrics. Most importantly, this framework provides the groundwork for implementing a learning healthcare system by providing evaluation criteria.

LIST OF ABBREVIATIONS

Not applicable.

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIALS

All data generated or analysed during this study are included in this published article [and its additional files]. Further information is available from the corresponding author.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

None

AUTHORS' CONTRIBUTIONS

CL and MO carried out the literature reviews, thematic analysis and writing the manuscript. RS was a major contributor in constructing the search strategy. CL designed the figures. LH and ML were major contributors to the research question and methodology. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Not applicable

LITERATURE LIST

1. Ng E, de Colombani P. Framework for Selecting Best Practices in Public Health: A Systematic Literature Review. J Public Health Res. 2015;4(3):577.
2. de Bruin J, Bos C, Struijs JN, Drewes HW, Baan CA. Conceptualizing learning health systems: A mapping review. Learn Health Syst. 2023;7(1):e10311.
3. Committee on the Learning Health Care System in America, Institute of Medicine. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
4. Boell SK, Cecez-Kecmanovic D. A hermeneutic approach for conducting literature reviews and literature searches. Communications of the Association for information Systems. 2014;34(1):12.
5. Hendricks-Sturrup RM, Cerminara KL, Lu CY. A Qualitative Study to Develop a Privacy and Nondiscrimination Best Practice Framework for Personalized Wellness Programs. J Pers Med. 2020;10(4).
6. Ford JH, 2nd, Gilson AM, Bryan G, Augustine C, Gassman M, Mott DA. Community pharmacy-based injectable naltrexone service delivery models and best practices. Res Social Adm Pharm. 2021;17(7):1332-41.
7. Yates J, Stanyon M, Challis D, Coleston-Shields DM, Dening T, Hoe J, et al. Developing a model of best practice for teams managing crisis in people with dementia: a consensus approach. BMC Psychiatry. 2020;20(1):505.
8. Naranjo A, Ojeda S, Giner M, Balcells-Oliver M, Canals L, Cancio JM, et al. Best Practice Framework of Fracture Liaison Services in Spain and their coordination with Primary Care. Arch Osteoporos. 2020;15(1):63.
9. Majmudar A, Jain AK, Chaudry J, Schwartz RW. High-performance teams and the physician leader: an overview. J Surg Educ. 2010;67(4):205-9.
10. Riley W, Davis SE, Miller KK, McCullough M. A model for developing high-reliability teams. J Nurs Manag. 2010;18(5):556-63.
11. Laur C, Keller HH. Implementing best practice in hospital multidisciplinary nutritional care: an example of using the knowledge-to-action process for a research program. J Multidiscip Healthc. 2015;8:463-72.
12. Rolin-Gilman C, Fournier B, Cleverley K. Implementing Best Practice Guidelines in Pain Assessment and Management on a Women's Psychiatric Inpatient Unit: Exploring Patients' Perceptions. Pain Manag Nurs. 2017;18(3):170-8.
13. Schippke J, Provvidenza C, Kingsnorth S. Peer support for families of children with complex needs: Development and dissemination of a best practice toolkit. Child Care Health Dev. 2017;43(6):823-30.
14. Wolfe DL, Hsieh JTC, Kras-Dupuis A, Riopelle RJ, Walia S, Guy S, Gillis K. An inclusive, online Delphi process for setting targets for best practice implementation for spinal cord injury. J Eval Clin Pract. 2019;25(2):290-9.
15. Younger CWE, Douglas C, Warren-Forward H. Medical imaging and informed consent - Can radiographers and patients agree upon a realistic best practice?

Radiography (Lond). 2018;24(3):204-10.
16. Findlay M, Rankin NM, Bauer J, Collett G, Shaw T, White K. “Completely and utterly flummoxed and out of my depth”: patient and caregiver experiences during and after treatment for head and neck cancer-a qualitative evaluation of barriers and facilitators to best-practice nutrition care. Support Care Cancer. 2020;28(12):5771-80.
17. Lorente-Granados MG, Quiñoz-Gallardo MD, Teixiné-Martín A, Arza-Alonso N, Suarez MD, Arias-Arias Á J. Implementation of the best practice guideline for pressure ulcer prevention: Development, results and sustainability. Enferm Clin (Engl Ed). 2020;30(3):198-211.
18. Taylor C, Mulligan K, McGraw C. Barriers and enablers to the implementation of evidence-based practice in pressure ulcer prevention and management in an integrated community care setting: A qualitative study informed by the theoretical domains framework. Health Soc Care Community. 2021;29(3):766-79.
19. Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, Tyrer P. Systematic review of involving patients in the planning and development of health care. Bmj. 2002;325(7375):1263.
20. Findlay M, Rankin NM, Shaw T, White K, Boyer M, Milross C, et al. Best Evidence to Best Practice: Implementing an Innovative Model of Nutrition Care for Patients with Head and Neck Cancer Improves Outcomes. Nutrients. 2020;12(5).
21. Zhao G, Kennedy C, Mabaya G, Okrainec K, Kiran T. Patient engagement in the development of best practices for transitions from hospital to home: a scoping review. BMJ Open. 2019;9(8):e029693.
22. Knox GE, Simpson KR, Garite TJ. High reliability perinatal units: an approach to the prevention of patient injury and medical malpractice claims. J Healthc Risk Manag. 1999;19(2):24-32.
23. Aufegger L, Shariq O, Bicknell C, Ashrafian H, Darzi A. Can shared leadership enhance clinical team management? A systematic review. Leadersh Health Serv (Bradf Engl). 2019;32(2):309-35.
24. Salas E, DiazGranados D, Weaver SJ, King H. Does team training work? Principles for health care. Acad Emerg Med. 2008;15(11):1002-9.
25. Nallamothu BK, Guetterman TC, Harrod M, Kellenberg JE, Lehrich JL, Kronick SL, et al. How Do Resuscitation Teams at Top-Performing Hospitals for In-Hospital Cardiac Arrest Succeed? A Qualitative Study. Circulation. 2018;138(2):154-63.
26. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients’ care. Lancet. 2003;362(9391):1225-30.
27. Briggs AM, Jordan JE, Jennings M, Speerin R, Bragge P, Chua J, et al. Supporting the Evaluation and Implementation of Musculoskeletal Models of Care: A Globally Informed Framework for Judging Readiness and Success. Arthritis Care Res (Hoboken). 2017;69(4):567-77.
28. Künzle B, Kolbe M, Grote G. Ensuring patient safety through effective leadership behaviour: a literature review. Safety science. 2010;48(1):1-17.
29. Johnsen BH, Westli HK, Espevik R, Wisborg T, Brattebø G. High-performing trauma teams: frequency of behavioral markers of a shared mental model displayed by team leaders and quality of medical performance. Scand J Trauma Resusc Emerg Med. 2017;25(1):109.

30. Kirk AP, McGlinsey A, Beckett A, Rudd P, Arbour R. Restraint Reduction, Restraint Elimination, and Best Practice: Role of the Clinical Nurse Specialist in Patient Safety. Clin Nurse Spec. 2015;29(6):321-8.
31. Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S, et al. Surgical team behaviors and patient outcomes. Am J Surg. 2009;197(5):678-85.
32. Lewis PS, Latney C. Achieve best practice with an evidence-based approach. Create a collaborative environment that improves patient care through consistent outcomes measurement. Nurs Manage. 2002;33(12):24, 6-8, 30.
33. Burgess L, Kynoch K, Hines S. Implementing best practice into the emergency department triage process. Int J Evid Based Healthc. 2019;17(1):27-35.
34. Feldstein J, Coussons H. Achieving robotic program best practice performance and cost versus laparoscopy: Two case studies define a framework for optimization. Int J Med Robot. 2020;16(3):e2098.
35. Brown TM, Garg S, Chandran AB, McNett M, Silverman SL, Hadker N. The impact of 'best-practice' patient care in fibromyalgia on practice economics. J Eval Clin Pract. 2012;18(4):793-8.
36. Fleiszer AR, Semenic SE, Ritchie JA, Richer MC, Denis JL. A unit-level perspective on the long-term sustainability of a nursing best practice guidelines program: An embedded multiple case study. Int J Nurs Stud. 2016;53:204-18.
37. Van Voorhis BJ, Thomas M, Surrey ES, Sparks A. What do consistently high-performing in vitro fertilization programs in the U.S. do? Fertil Steril. 2010;94(4):1346-9.
38. Higuchi KS, Davies BL, Edwards N, Ploeg J, Virani T. Implementation of clinical guidelines for adults with asthma and diabetes: a three-year follow-up evaluation of nursing care. J Clin Nurs. 2011;20(9-10):1329-38.
39. Sharplin G, Adelson P, Kennedy K, Williams N, Hewlett R, Wood J, et al. Establishing and Sustaining a Culture of Evidence-Based Practice: An Evaluation of Barriers and Facilitators to Implementing the Best Practice Spotlight Organization Program in the Australian Healthcare Context. Healthcare (Basel). 2019;7(4).
40. Guzman G, Fitzgerald JA, Fulop L, Hayes K, Poropat A, Avery M, et al. How best practices are copied, transferred, or translated between health care facilities: A conceptual framework. Health Care Manage Rev. 2015;40(3):193-202.
41. Vinson AH. Culture as infrastructure in learning health systems. Learn Health Syst. 2021;5(3):e10267.
42. Wouters RHP, van der Graaf R, Voest EE, Bredenoord AL. Learning health care systems: Highly needed but challenging. Learn Health Syst. 2020;4(3):e10211.
43. Braithwaite J, Herkes J, Ludlow K, Testa L, Lamprell G. Association between organisational and workplace cultures, and patient outcomes: systematic review. BMJ Open. 2017;7(11):e017708.
44. DiCuccio MH. The Relationship Between Patient Safety Culture and Patient Outcomes: A Systematic Review. J Patient Saf. 2015;11(3):135-42.
45. Khan MF, Sewell MD, Alrawi A, Taif S, Divani K. Can a culture of team psychological safety and MDT proforma improve team performance and patient outcomes in spinal MDTs? Br J Neurosurg. 2022:1-5.
46. Thomas L, Galla C. Republished: Building a culture of safety through team training and engagement. Postgrad Med J. 2013;89(1053):394-401.

47. Yuce TK, Yang AD, Johnson JK, Odell DD, Love R, Kreutzer L, et al. Association Between Implementing Comprehensive Learning Collaborative Strategies in a Statewide Collaborative and Changes in Hospital Safety Culture. JAMA Surg. 2020;155(10):934-40.
48. Pavithra A. Towards developing a comprehensive conceptual understanding of positive hospital culture and approaches to healthcare organisational culture change in Australia. J Health Organ Manag. 2021;ahead-of-print(ahead-of-print).
49. Ananthakrishnan AN, Hoffmann RG, Saeian K. Higher physician density is associated with lower incidence of late-stage colorectal cancer. J Gen Intern Med. 2010;25(11):1164-71.
50. Farahani M, Price N, El-Halabi S, Mlaudzi N, Keapoletswe K, Lebelonyane R, et al. Impact of Health System Inputs on Health Outcome: A Multilevel Longitudinal Analysis of Botswana National Antiretroviral Program (2002-2013). PLoS One. 2016;11(8):e0160206.
51. Fleisher JM, Lou JQ, Farrell M. Relationship between physician supply and breast cancer survival: a geographic approach. J Community Health. 2008;33(4):179-82.
52. Gulliford MC. Availability of primary care doctors and population health in England: is there an association? J Public Health Med. 2002;24(4):252-4.
53. Sakai R, Fink G, Kumamaru H, Kawachi I. The Impact of Pediatrician Supply on Child Health Outcomes: Longitudinal Evidence from Japan. Health Serv Res. 2016;51(2):530-49.
54. Van Bulck L, Goossens E, Luyckx K, Apers S, Oechslin E, Thomet C, et al. Healthcare system inputs and patient-reported outcomes: a study in adults with congenital heart defect from 15 countries. BMC Health Serv Res. 2020;20(1):496.
55. Eriksson CO, Stoner RC, Eden KB, Newgard CD, Guise JM. The Association Between Hospital Capacity Strain and Inpatient Outcomes in Highly Developed Countries: A Systematic Review. J Gen Intern Med. 2017;32(6):686-96.
56. Lentz CM, De Lind Van Wijngaarden RAF, Willeboordse F, Hooft L, van der Laan MJ. Dedicated teams to optimize quality and safety of surgery: A systematic review. Int J Qual Health Care. 2022;34(4).
57. Ricci MA, Brumsted JR. Crew resource management: using aviation techniques to improve operating room safety. Aviat Space Environ Med. 2012;83(4):441-4.
58. Roberts NK, Williams RG, Schwind CJ, Sutyak JA, McDowell C, Griffen D, et al. The impact of brief team communication, leadership and team behavior training on ad hoc team performance in trauma care settings. Am J Surg. 2014;207(2):170-8.
59. McGinnis JM, Fineberg HV, Dzau VJ. Advancing the Learning Health System. N Engl J Med. 2021;385(1):1-5.
60. Teisberg E, Wallace S, O'Hara S. Defining and Implementing Value-Based Health Care: A Strategic Framework. Acad Med. 2020;95(5):682-5.
61. Scott T, Mannion R, Davies H, Marshall M. The quantitative measurement of organizational culture in health care: a review of the available instruments. Health Serv Res. 2003;38(3):923-45.

CHAPTER 3
Additional files

ADDITIONAL FILE 1

MEDLINE AND EMBASE SEARCH STRATEGY

Database: Medline (OVID)

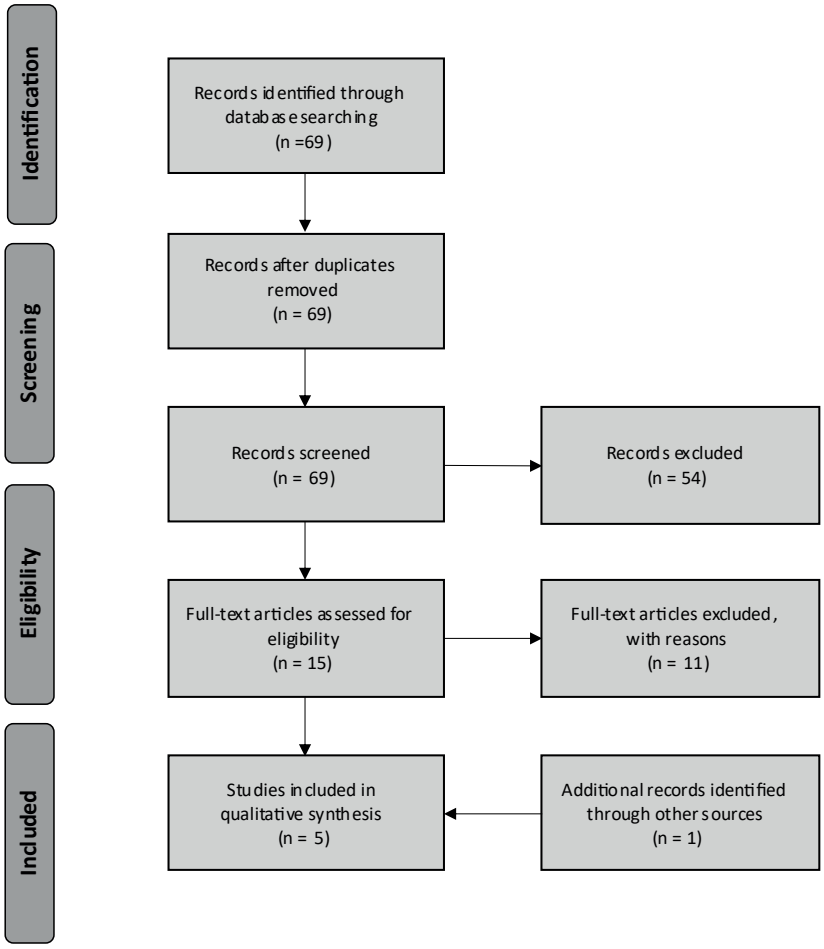
Search terms:

1: (theor*[ti] or concept*[ti] or framework*[ti] or model*[ti]) AND (best-practice*[ti])

2: limit 1 to (English language and yr="2018-Current")

ADDITIONAL FILE 2

PRISMA FLOWCHART



ADDITIONAL FILE 3

GLOSSARY

Term	Definition
Culture of trust	The shared notion that members of a team will perform their expected tasks and look out for the needs of the team.[24]
Supporting behavior	Team members show support for each other and can anticipate each other's needs, supporting behaviors include support through teaching and coaching, supervision, monitoring and concern for team member's well-being.[23, 24]
Closed loop communication	Communication that is actively acknowledged by the receiver and actively affirmed by the send, often verbally.[10, 24]
Situational awareness	The ability to observe, interpret and communicate the level of urgency of a given situation in relation to a patient's health underpinned by experience and training.[10, 23]
Shared leadership	A distribution of leadership across the team, this type of leadership is more dynamic and interactive than the traditional leadership style with one appointed leader.[23]
Shared mental model	The team has a shared understanding of the problem and the plan; all members are involved in this understanding. Clear roles and responsibilities are essential for this shared understanding. [10, 23, 24]
Dedicated teams	Teams dedicated to a certain clinical task, often with certain skills needed for the task, dedicated resuscitation teams are an example.[25]
Coordination of service	Coordination of the team involves planning, preparation, delegation and directing.[23]
Interdependence	Interdependence promotes efficient teamwork; the idea is that mutual dependence regarding specific skills within a team limits duplication of work and therefore promotes efficient use of resources.[9]
Technical skills	Practical skills contributing to providing optimum patient care.[10]
Non-technical skills	Intrapersonal and interpersonal skills contributing to providing optimum patient care, communication and teamwork are important examples in healthcare.[10]
Models of care	Models of care are a means to achieve change in clinical practice, models of care translate evidence-based policies into principles that can be implemented in practice.[20, 27]
Culture of safety	The shared beliefs and values of the individuals compromising an organization that patient safety is a priority, resulting behaviors to continually minimize patient harm.[10]
Quality indicators	Clinical and/or operational outcomes used to benchmark quality of care.[20, 32]
Culture of shared accountability	All team members share accountability in providing optimum healthcare.[34, 36]
Reflexivity	Team members are actively involved in the evaluation process of a best practice.[20]

ADDITIONAL FILE 4

TOOL FOR THE EVALUATION OF A BEST PRACTICE

TEMPLATE FOR THE DOCUMENTATION FOR THE EVALUATION OF A BEST PRACTICE IN HEALTHCARE

The following tool is a template for conducting a detailed documentation for the evaluation of a best practice in healthcare. It focuses on documentation to aid evaluation of an already identified best practice in a hospital setting.

Evidence sources include evaluations, clinical reports and personal experiences.

Section 1: General information
Describe the best practice being documented, be as concrete and specific as possible. (specialty/ sub-specialty)
Organization being documented
Location (country/ region)
Date of documentation
Contact person (function, email address, phone number)

Section 2a: Contextual information *Relevant to needs*
Describe the target group of the best practice
Identify all stakeholders relevant and important to the best practice

Document how the best practice is relevant to the needs of the:
a. Target group

b. Stakeholders

Was stakeholder consensus reached regarding the following aspects?

☐ Development of the best practice

☐ Implementation of the best practice

☐ Other:

Document the organizational support the best practice receives and whether this meets the needs.
Organizational support is a broad concept and can include things such as access to necessary expertise, information technology and financial resources.

Section 2b: Contextual information
Relevant to setting

Describe the culture in which the practice is implemented

Document the resources available to the practice
Resources can include but are not limited to financial resources, human resources, information resources, skills, training and education.

Describe the workload of the staff involved in delivering the practice
Is the workload experienced as low/ acceptable/ high?

Section 3: Processual information

Do stakeholders participate in the following activities?

☐ Decision making processes
If yes, please describe through which activities they participate:

☐ Evaluation of the best practice
If yes, please describe through which activities they participate:

Document the team composition of the team responsible for providing patient care in this practice.

Please indicate on the scale below to what extent the following characteristics regarding the team culture are present in this practice:

Collegial rather than hierarchical behavior

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Supporting behavior

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

A clear service purpose

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Document whether the following characteristics regarding communication are present in this practice:

☐ Open communication

☐ Closed loop communication

☐ Clear/ standardized language

☐ Interdisciplinary communication

☐ Situational awareness

☐ Other:

Document whether the following characteristics regarding leadership are present in this practice:

☐ Shared leadership

☐ Leadership support

☐ Decision making

☐ Other:

To what extent do you agree with the following statement? Please indicate on the scale below.
"There is a sense of unity present within the team"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Please elaborate why there is (not) a sense of unity within the team?
Document what contributes to the presence/ absence of a sense of unity within the team.

Document whether the following characteristics regarding the organizational structure are present in this practice:

☐ Dedicated teams

☐ Clear roles and responsibilities

☐ Coordination of services

☐ Interdependence

☐ Other:

Describe the training and education staff receive including the frequency thereof:

Describe the staff competency relevant to the best practice:

Describe staff skills in terms of:
a. Technical skills
Technical skills are practical skills contributing to provision of optimum patient care.

b. Non-technical skills (this may include skills such as interpersonal skills)

Document whether the following characteristics regarding the use of best available evidence are present in this practice:

☐ The best available evidence is used in the practice

☐ Best available evidence is adhered to

☐ Resources are available allowing the use of the best available evidence

☐ Scientific evidence supports the practice

☐ Awareness exists regarding evidence-based guidelines

☐ Models of care are utilized

☐ Other:

To what extent do you agree with the following statement? Please indicate on the scale below.
"There is a culture of improvement present within the practice"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Please elaborate why there is (not) a culture of improvement?
Document what contributes to the presence/ absence of a culture of improvement.

Document whether the following characteristics regarding safety management are present in this practice:

☐ Safety is managed proactively

☐ Leader provides updates to the team

☐ Injuries and near-misses are detected and analyzed

☐ Processes are designed for safety

☐ Patient safety is a priority

☐ Communication is proactive (this may include handovers at shift changes)

☐ Emergencies are rehearsed during team trainings

☐ Other:

To what extent do you agree with the following statement? Please indicate on the scale below.
"There is a culture of safety present within the practice"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Please elaborate why there is (not) a culture of safety?
Document what contributes to the presence/ absence of a culture of safety.

Section 4: Outcomes

Describe how this practice was identified as a best practice, include which outcomes were used to reach this conclusion.

Document the clinical outcomes of this practice:
a. Which clinical parameters are monitored and what were the outcomes?

b. Is patient satisfaction monitored? If yes, how and what was the level of satisfaction?

c. How is quality of care measured? (for instance through the use of certain quality indicators)

d. How is the performance of the practice measured and provide performance results? (for instance through the use of clinical reports)

Document the efficiency of the practice
Efficiency can include several performance metrics such as surgical time and turnover time, which measures are relevant is often practice dependent. Please document the ones deemed relevant in this practice.

Document the use and incorporation of information technology to monitor efficiency.

Document the cost-effectiveness of the practice.

Document whether the practice sets goals and what these entail.
Goals may be set through an action plan of long and short-term goals for instance. The types of goals may entail changing processes, structure or improving outcomes.

Document whether the following resources are sustainable in this practice:

☐ Financial resources

☐ Staffing

☐ Activities (such as education, reminders, evaluation)

☐ Training of new staff members

☐ Availability of educational materials

☐ Other:

Is continuity embedded in the program delivery?

☐ Yes ☐ No

If yes, please elaborate how this continuity is embedded in delivery?

Document whether the following characteristics regarding the sustainability of culture are present in this practice:

☐ The best practice is perceived as advantageous by those participating in the best practice

☐ Hospital staff were incorporated in implementing the best practice

☐ The leadership is sustainable

☐ There is a culture of shared accountability

☐ The organization is committed to the best practice

☐ Other:

Section 5: Learning capacity

Document whether the following characteristics regarding the evaluation of performance are present in this practice:

☐ Performance is monitored and evaluated

☐ Performance indicators are used to monitor quality of care

☐ Adverse events are debriefed

☐ The team is reflexive

☐ Other:

Are there quality improvement initiatives?

☐ Yes ☐ No

If yes, please elaborate what these initiatives are and if these are continuous of nature?

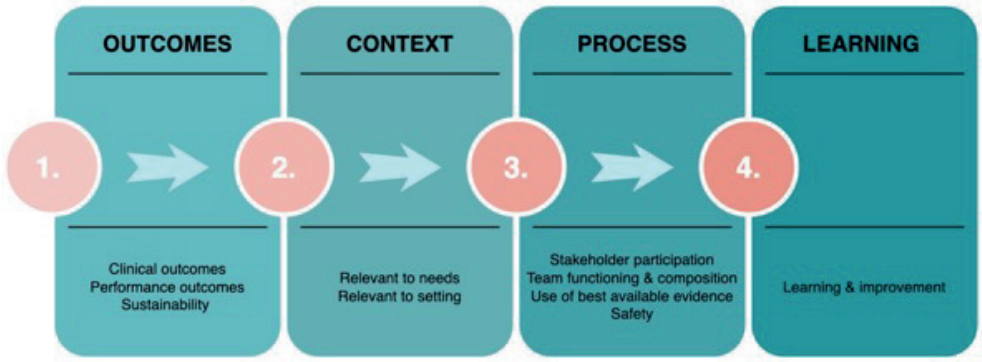
To what extent do you agree with the following statement? Please indicate on the scale below.
"There is a culture of learning present within the practice"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Please elaborate what constitutes to and why there is (not) a learning culture present? Document what contributes to the presence/ absence of a learning culture.

ADDITIONAL FILE 5

FLOWCHART OF A PROPOSED EVALUATION OF A BEST PRACTICE



Flowchart of a proposed evaluation of a best practice. *This flowchart depicts the steps of a proposed evaluation process of a best practice. The first step being the identification of a best practice through determining a practice with most favorable outcomes. The next steps of the evaluation process include assessing the context, process and learning capacity of the practice, the framework in this study highlights the quality indicators that play a role in a best practice.*

CHAPTER 4

Designing tailored maintenance strategies for systematic reviews and clinical practice guidelines using the Portfolio Maintenance by Test-Treatment (POMBYTT) framework

M.S. Oerbekke
R.G. Elbers
M.J. van der Laan
L. Hooft

BMC Med Res Methodol. 2024 Feb 2;24(1):29
doi: 10.1186/s12874-024-02155-z
PMID: 38308228

ABSTRACT

BACKGROUND

Organizations face diverse contexts and requirements when updating and maintaining their portfolio, or pool, of systematic reviews or clinical practice guidelines they need to manage. We aimed to develop a comprehensive, theoretical framework that might enable the design and tailoring of maintenance strategies for portfolios containing systematic reviews and guidelines.

METHODS

We employed a conceptual approach combined with a literature review. Components of the diagnostic test-treatment pathway used in clinical healthcare were transferred to develop a framework specifically for systematic review and guideline portfolio maintenance strategies.

RESULTS

We developed the Portfolio Maintenance by Test-Treatment (POMBYTT) framework comprising diagnosis, staging, management, and monitoring components. To illustrate the framework's components and their elements, we provided examples from both a clinical healthcare test-treatment pathway and a clinical practice guideline maintenance scenario. Additionally, our literature review provided possible examples for the elements in the framework, such as detection variables, detection tests, and detection thresholds. We furthermore provide three example strategies using the framework, of which one was based on living recommendations strategies.

CONCLUSIONS

The developed framework might support the design of maintenance strategies that could contain multiple options besides updating to manage a portfolio (e.g. withdrawing and archiving), even in the absence of the target condition. By making different choices for variables, tests, test protocols, indications, management options, and monitoring, organizations might tailor their maintenance strategy to suit specific contexts and needs. The framework's elements could potentially aid in the design by being explicit about the operational aspects of maintenance strategies. This might also be helpful for end-users and other stakeholders of systematic reviews and clinical practice guidelines.

Keywords: clinical practice guidelines as topic, systematic reviews as topic, concept formation, theoretical model, need for updating, maintenance

1. BACKGROUND

Fifteen percent of the systematic reviews (SRs)[1] and eight percent of the recommendations in clinical practice guidelines (CPGs)[2] may be out of date within the first year after their publication. Over time, there could be changes in the evidence on the harms, benefits, and availability of interventions, and changes in important outcomes for instance.[3] Neglecting such changes could cause SR conclusions and CPG recommendations to become invalid, potentially leaving clinical practice sub-optimal. Updating thus seems a reasonable option to manage outdated SRs and CPGs. The problem of when and how to update SRs was highlighted more than one decade ago [4] and more than two decades ago for CPGs.[3] The Cochrane Collaboration provides guidance on when and how to update an SR.[5,6] Furthermore, specific strategies to detect the need for updating were being developed for SRs, such as the Ottawa[7] and RAND methods,[8] and for CPGs.[3,9-12] Previous published systematic reviews provided overviews of such methods for both SRs [4] and CPGs.[9,13]

A large variety of strategies to assess when to update SRs or CPGs can be observed in the literature.[13,14] Even within similar assessments, such as literature searches to identify new evidence, there is a variety in how the assessment is performed. For example, search strategies can be limited to specific journals[7,8,10] and publication type.[7,10] The full search strategy of the original reviews can be updated,[15] additional searches can be performed in a guideline database,[10] experts can be consulted,[3,12] or studies can be tracked in trial registries.[15,16] New strategies and insights about updating strategies are still being introduced, such as the concept of living SRs[17] and living CPG recommendations.[18] Current strategies may not be useful for the context, capabilities, or the needs of all organizations performing updates. Different choices can be made in designing strategies to accommodate for the different contexts, capabilities, and needs. For example, strategies with an extensive literature search for each key question could be too resource intensive for CPG developing organizations managing a large portfolio (i.e. a pool of SRs or CPGs that is managed by the organization). Such considerations might prevent adoption or cause revisitation of existing strategies and could partially explain why new strategies are still being reported. Cochrane, for example, has changed their updating principles on several occasions reflecting their experience that they were not yet able to constantly keep their entire portfolio of SRs up-to-date over time.[15] Furthermore, updating might not be the only option available to manage an outdated SR or CPG. Withdrawal or archiving could be suitable alternative options to maintain the portfolio of SRs or CPGs as well, where withdrawal completely removes the SR or CPG from the portfolio and archiving still allows end-users to access the information while no longer actively maintained. It seems, rather, that there could be a need for guidance to design and tailor maintenance strategies instead of updating strategies.

A framework with explicit underlying key components and elements for designing portfolio maintenance strategies appears to be missing at present. A new framework therefore should identify and explain these key components and elements in the context of a maintenance strategy, potentially enabling organizations to tailor a strategy according to their context, capabilities, needs and available resources. We aimed to develop and describe such a theoretical framework for designing and tailoring maintenance strategies for managing

portfolios of SRs and CPGs.

2. METHODS

A literature review was conducted to gain a comprehensive overview of considerations, signals, or indicators for updating SRs and CPGs. The literature review (methodology reported in Additional file 1) is not exhaustive, as we did not need to capture all data on every domain. During the data-extraction we observed that other management options were available besides (not) updating. For example, withdrawing an SR or CPG. While exploring the extracted data thereafter, we observed a supposed interrelatedness between some considerations, signals, and indicators. Through discussion among the authors, we believed that the interrelatedness and the availability of multiple management options had an analogy to a diagnostic test-treatment pathway in the clinical care setting. In a test-treatment pathway, medical tests are linked to management actions through pathways so that test results guide clinical management.[19] We envisioned a parallel scenario where considerations, signals, and indicators guide the selection of appropriate management actions for SR and CPG maintenance. We therefore transferred the diagnosis, staging, management, and monitoring concepts of a diagnostic test-treatment pathway to develop a theoretical framework for designing and tailoring SR and CPG maintenance strategies. We recognize that alternative conceptual frameworks or constructs could have been considered as well, however the analogy to a diagnostic test-treatment pathway resonated with us due to its apparent suitability to represent how considerations, signals, and indicators could be linked to management. The extracted data from our literature review were qualitatively analyzed and these results were used to provide some possible examples of key elements in the framework. Thus, data from the literature review both directed us to use a diagnostic test-treatment strategy analogy and provided examples for the framework's elements. To explicitly clarify the components and elements in the framework, we describe both a clinical healthcare example and a CPG maintenance scenario. The 2018 European Society of Cardiology and European Society of Hypertension guideline for the management of arterial hypertension was used as clinical example.[20] The CPG maintenance strategy scenario was based on considerations and signals found in the literature review, however, modified for illustrative purposes. Tables concerning the clinical example and the CPG maintenance scenario represent subsequent steps in the diagnostic test-treatment pathway. Results from our literature review were mapped at our own discretion to the specific test-treatment components of the maintenance strategy to provide examples, even though the extracted data may have been described for other purposes in the original references.

3. RESULTS

3.1 LITERATURE SEARCH

Fifty-four references were included. The study selection flow (Figure A1 in Additional file 1) and reasons for exclusion of full-text references are reported in Additional file 1 (Table A1). General characteristics of the included studies are described in Additional file 1 (Table A2). Results from the literature review are provided as possible examples for elements in the framework in Additional File 1 (Tables A3 to A9).

3.2 A THEORETICAL FRAMEWORK FOR PORTFOLIO MAINTENANCE STRATEGIES

The Portfolio Maintenance by Test-Treatment (POMBYTT) framework is shown in Figure 1. The theoretical POMBYTT framework is intended to help design and tailor maintenance strategies for portfolios consisting of SRs or CPGs. Components of a diagnostic test-treatment pathway are transferred to a portfolio maintenance context: diagnosis, staging, management, and monitoring (Table 1). These concepts in the framework are outlined in Additional file 1 (Figure A2). Specific terminology is used throughout the description of the framework and a glossary of terms can be found in Table 2.

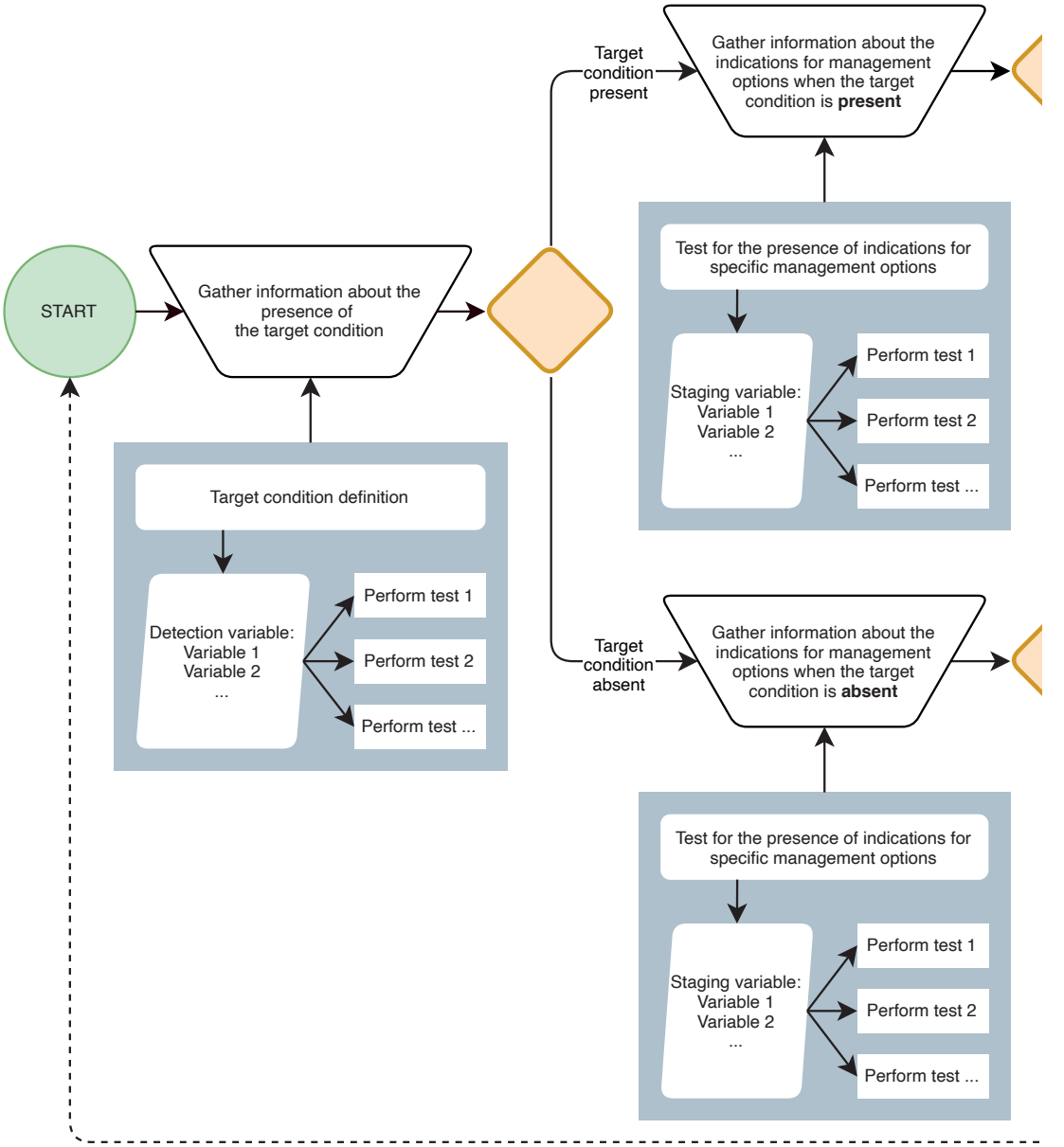
3.2.1 DIAGNOSIS

The target condition must be defined before it can be detected with diagnostic tests. Let's consider the example of determining whether a CPG recommendation is outdated. In this care we can define a recommendation as outdated when at least one new relevant peer-reviewed article is published after the previous search date. It is important to have a specific definition that outlines the unit of analysis. In the context of SR or CPG maintenance strategies, the unit of analysis can be the entire SR or CPG, or it can focus on the SR conclusion or CPG recommendation. Like diagnosing a medical condition in clinical practice, we need one or more detection variables (Table A3 in Additional file 1) that provide information about the presence or absence of the target condition. In the provided example in Table 3, the detection variable was "*new available evidence*" but it is worth noting that other detection variables can be used depending on the specific context. To measure these detection variables, we can use detection tests (Table A4 in Additional file 1). For example, a literature search in a database like MEDLINE can be used as a test to measure the detection variable "*new available evidence*". The test protocol for the literature search can vary, including the choice of using multiple databases, limiting the search to specific databases, or even limiting to a few specific journals. Additionally, literature selections can be performed by a single person or in a double-blind fashion and any selection procedure in between.

A detection test threshold (Table A5 in Additional file 1) is used to determine whether it is likely that the target condition is present or not. The threshold determines how the target condition is defined. In Table 3, the threshold to detect the target condition was any new relevant peer-reviewed article (i.e. ≥1). If the threshold was increased to at least 3 new relevant peer-reviewed articles a different definition of the target condition is detected (i.e. outdated when ≥3 new relevant articles). See Table 3 for a clinical example and a CPG scenario.

3.2.2 STAGING

The staging process occurs after determining whether the target condition is present or


START
Gather information about the presence of the target condition
Target condition definition
Detection variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...
Target condition present
Gather information about the indications for management options when the target condition is **present**
Test for the presence of indications for specific management options
Staging variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...
Target condition absent
Gather information about the indications for management options when the target condition is **absent**
Test for the presence of indications for specific management options
Staging variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...

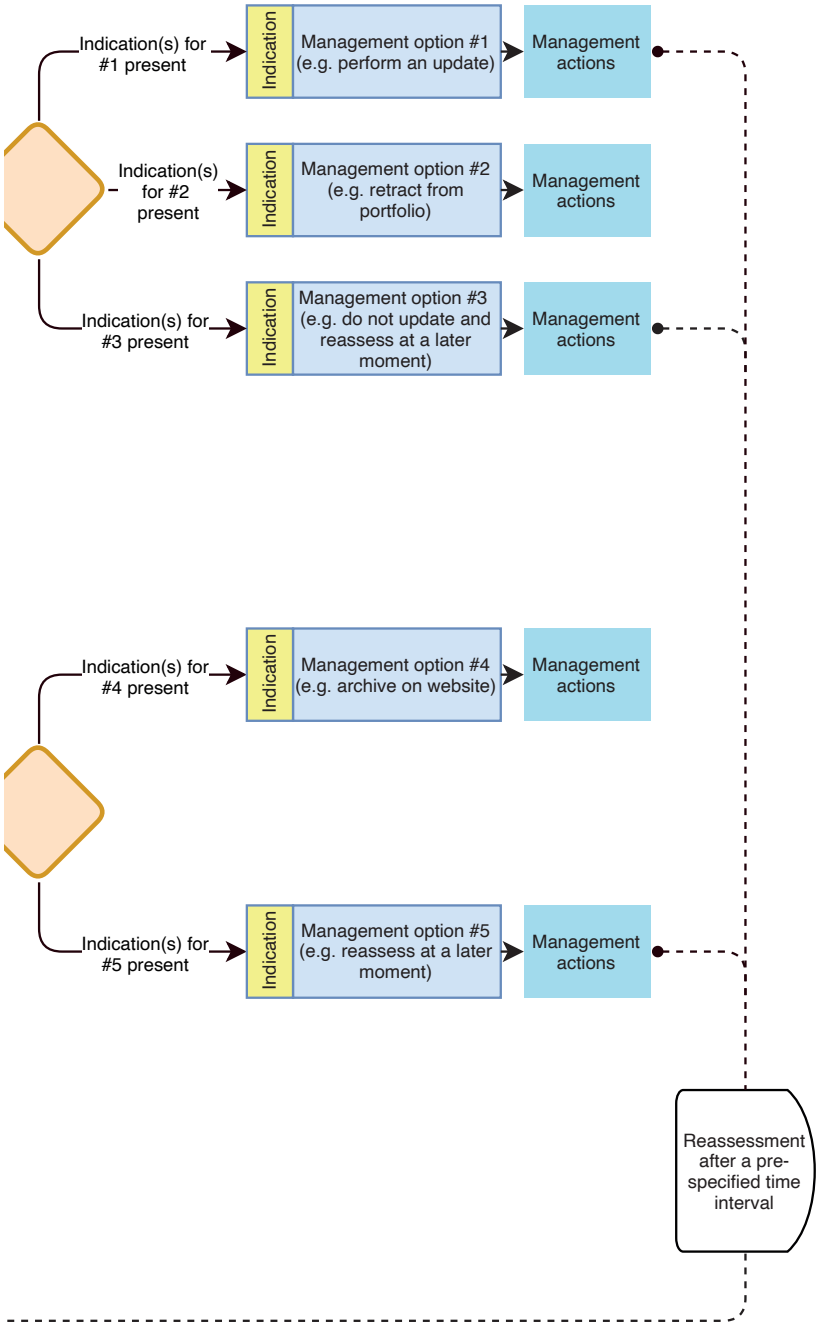


Figure 1. The Portfolio Maintenance by Test-Treatment framework. *The figure shows the framework depicted as a flow diagram in analogy to a diagnostic test-treatment pathway. Tests are performed (grey boxes, not outlined), choices are made (outlined orange diamonds), management options (outlined blue boxes) are selected based on indications (outlined yellow boxes), subsequent management actions are performed (blue boxes, not outlined), and predefined time intervals are used for reassessments (dashed line).*

Table 1 – An overview of test-treatment pathway components compared to an SR or CPG portfolio maintenance strategy.

Component	Diagnostic test-treatment pathway	SR or CPG portfolio maintenance strategy
Diagnosis *(see section 3.2.1)*	A "prevalent disease or condition" is diagnosed. Diagnosis requires a clear definition of the target condition (i.e. the disease or condition) and the use of tests to assess the likelihood of its presence or absence. In a clinical scenario, diagnostic tests include anamnesis (medical history) along with other clinical tests, such as physical examination or imaging. By setting a threshold within the test, healthcare professionals can determine the likelihood of the disease being present.	The diagnosis of "prevalent outdatedness" involves the use of tests to determine the target condition's likelihood of presence or absence. Similar to a clinical test-treatment pathway, these tests help to determine whether the target condition is likely to be present or absent. It is important to define outdatedness in advance. A threshold within a test can then be established to assess whether outdatedness is likely to be present or absent.
Staging *(see section 3.2.2.)*	Staging of a disease or condition is conducted to evaluate its stage or degree of severity. Information necessary for staging is gained with tests. Within these tests, thresholds are used to indicate the severity of the disease or condition.	When "prevalent outdatedness" is diagnosed, its severity is assessed using tests. Several stages or degrees of outdatedness might result in different decisions regarding appropriate management options. Thresholds within the tests serve as criteria for categorizing stages or severity of outdatedness. Staging can also be performed when the systematic review (conclusion) or clinical practice guideline (recommendation) is still op to date.
Management *(see section 3.2.3)*	There are specific indications to select management options. Staging provides valuable information to determine which management option is most appropriate. Once a management option is chosen, actions are performed to initiate and carry out the appropriate care.	Management options, such as updating, not updating, or withdrawing the SR or CPG have specific indications. Staging provided information to decide which management option is appropriate. Additionally, organizations may have several other management options (e.g. archive, re-endorse). Once the management option is selected, actions are performed specific for that option. For example, an update of the SR or CPG is initiated and carried out.
Monitoring *(see section 3.2.4)*	Patients are periodically seen during follow-up visits as part of their ongoing care. The purpose of these visits is to monitor the disease or condition over time and evaluate the success of the chosen management option, detect signs of disease recurrence, or identify any disease progression.	Periodic evaluation of SRs or CPGs can be performed to assess whether there is an occurrence, recurrence, or progression of outdatedness. This periodic evaluation is especially relevant when the initial staging of the target condition did not meet the threshold for selecting a specific management option (e.g. updating).

CPG: Clinical Practice Guideline
SR: Systematic review

Table 2 – Glossary of terms used in the conceptual maintenance strategy for systematic reviews (SRs) and clinical practice guidelines (CPGs)

Component	Element	Description as used in the maintenance of SRs and CPGs
Diagnosis	*Target condition*	The predefined condition of the SR (conclusion) or CPG (recommendation) that is to be detected by one or multiple tests.
	Detection variable	A variable or characteristic of the predefined target condition on which a detection test specifically measures information.
	Detection test	A test to measure or obtain information on a detection variable to determine whether the target condition is likely to be present or absent.
	Detection test protocol	The protocol or manner how the detection test is carried out to measure the detection variable or obtain information.
	Detection test threshold	A predefined threshold for the detection variable in a detection test to define the presence or absence of the target condition.
Staging	*Staging variable*	A variable or characteristic of the stage or severity of the target condition on which a staging test specifically measures information. A staging variable can also concern a status or circumstance not related to the target condition itself.
	Staging test	A test to measure or obtain information on a staging variable to determine the stage or severity of the target condition. A staging test may also capture a status or circumstance not related to the target condition itself.
	Staging test protocol	The protocol or manner how the detection test is carried out to measure the staging variable or obtain information.
	Staging test threshold	One or more predefined thresholds for the staging variable in a staging test to define the stage, severity, status, or circumstance.
Management	*Management indication*	A specific stage, severity, status, or circumstance that guides the decision to the appropriate management option.
	Management options	A predefined approach to handle an SR (conclusion) or CPG (recommendation) when the target condition is present or absent, and depending on the specific stage, severity, status, or circumstance. The choice for a management option is guided by its indication(s).
	Management actions	Actions that follow from the decision for a specific management option. The actions performed to carry out the chosen management of the SR (conclusion) or CPG (recommendation).
Monitoring	*Recurrence*	The recurrence of the predefined target condition after an update. Monitoring is used to detect recurrences.
	Progression	The progression of the stage, status, or severity of the predefined target condition. Monitoring is used to detect progression until a threshold for a specific management option is reached (e.g. update).

CPG: Clinical Practice Guideline
SR: Systematic review

Table 3 – Example of the target condition and diagnostic test in a clinical example and CPG scenario. The CPG scenario is based on considerations and signals found in our literature search, however, modified for illustrative purposes.

	Clinical healthcare example	GPG maintenance scenario
Objective	To detect whether there is arterial hypertension	To detect whether a recommendation is outdated
Target condition definition	We consider high blood pressure to be prevalent when there is at least 140 mmHg systolic arterial pressure over 90 mmHg diastolic arterial pressure.	We consider the recommendation to be outdated when at least one new relevant peer-reviewed article is published after the previous search date.
Detection variable	Pressure on artery walls	New available evidence
Detection test	Auscultatory sphygmomanometer	Literature search in MEDLINE
Detection test protocol	Three blood pressure measurements are taken one to two minutes apart after the patient is seated in a quiet environment for five minutes. The last two blood pressure measurements are averaged.	A sensitive search string for MEDLINE is constructed. Search results are screened on title and abstracts by two independent assessors. Conflicts are resolved by a third independent assessor. The resulting potentially relevant articles are read and selected by two assessors independently based on the full text. A third assessor resolves any conflict in the full text selection.
Detection test threshold	140 mmHg systolic arterial pressure over 90 mmHg diastolic arterial pressure	Any new peer-reviewed scientific article published after the previous search date

absent (Figure A2 in Additional file 1). The goal is to gain information about the severity, status, or stage. This is done by utilizing one or multiple staging variables (Table A6 in Additional file 1), staging tests (Table A7 in Additional file 1), and staging thresholds (Table A8 in Additional file 1).

Staging tests are used to measure information on the staging variable. Staging thresholds are defined in order to define the different stages or severity. The information obtained from the staging tests, along with the staging thresholds, guide the decision-making process towards an appropriate management option. Identical to detection tests, staging tests have variations in the test protocol and changing the thresholds also changes the definition of the stage, status, or severity. Table 4 provides a clinical example and a CPG scenario offering an understanding of the staging process.

It can still be important to perform staging tests when the target condition is absent, as several management options might still be available (see Table 5). A specific status or circumstance may be present that guides the management decision towards a specific management option.

3.2.3 MANAGEMENT

A management option is chosen once the severity, stage, or status is reasonably determined. Multiple management options can be available besides just updating an outdated SR or

Table 4 – Example of a single staging test and its thresholds in the presence of the target condition. The CPG scenario is based on considerations and signals found in our literature search, however, modified for illustrative purposes.

	Clinical healthcare example	GPG maintenance scenario
Target condition	Arterial hypertension is present (≥140/90 mmHg)	The recommendation is outdated (There was new peer-reviewed scientific evidence published)
Objective	To detect the severity, stage, or status of the prevalent hypertension so that an appropriate management option can be selected	To detect the severity, stage, or status of the prevalent outdatedness so that an appropriate management option can be selected
Staging variable	Magnitude of pressure on artery walls in a free-living setting	Likelihood of potential changes in the strength of the recommendation
Staging test	Ambulatory blood pressure monitoring	Survey among experts
Staging test protocol	The patient receives a blood pressure measuring device to wear over the course of 24 hours. The device is programmed to record the blood pressure once every 30 minutes. Blood pressure measurements are averaged for daytime and nighttime.	A survey among experts is performed by using the results from the literature selection. Experts are given the current recommendation and literature, while being asked whether the newly identified literature is likely to change the recommendation's strength requiring a dichotomous answer (yes/no).
Staging test thresholds	**Grade 1:** 140-159 mmHg systolic and/or 90-99 mmHg diastolic arterial pressure **Grade 2:** 160-179 mmHg systolic and/or 100-109 mmHg diastolic arterial pressure **Grade 3:** ≥180 mmHg systolic and/or ≥110 mmHg diastolic arterial pressure	**Very unlikely:** 0-20% of the surveyed experts indicated that the new evidence is likely to change the strength of the recommendation when updated. **Reasonably unlikely:** 20-40% of the surveyed experts indicated that the new evidence is likely to change the strength of the recommendation when updated. **Unclear:** 40-60% of the surveyed experts indicated that the new evidence is likely to change the strength of the recommendation when updated. **Reasonable likely:** 60-80% of the surveyed experts indicated that the new evidence is likely to change the strength of the recommendation when updated. **Very likely:** 80-100% of the surveyed experts indicated that the new evidence is likely to change the strength of the recommendation when updated.

CPG. Such options can include withdrawal, archiving, choosing not to update, or deferring an update to a later time. Similarly, when the target condition is not present, there can be multiple management options available as well (Table 5).

For example, if certain indicators are met, such as the CPG recommendation being fully implemented and there is minimal practice variation, it may be appropriate to archive the SR or CPG. Each management option has its own specific indications (Table A9 and Figure A2 in Additional file 1). The presence or absence of these indications, as evaluated using staging tests, guide the decision for specific management options. This process is similar to selecting appropriate management in clinical practice (see Table 6).

Once a management option is chosen, subsequent actions are undertaken to carry out the management option. These actions can be described in detail and can usually be found

Table 5 – Example of staging and management when the target condition is absent.

		Clinical healthcare example	GPG maintenance scenario
Target condition		Arterial hypertension is **absent** (<140/90 mmHg)	The recommendation is **not outdated** (There was no new peer-reviewed scientific evidence published)
Staging	*Staging variable*	Magnitude of pressure on artery walls in a free-living setting	Practice variation
	Staging test	Ambulatory blood pressure monitoring	Data registry analysis
	Staging test protocol	The patient receives a blood pressure measuring device to wear over the course of 24 hours. The device is programmed to record the blood pressure each 30 minutes. Blood pressure measurements are averaged for daytime and nighttime.	Data from registries are obtained. Variables concerning the CPG recommendation are analyzed in statistical software to show whether there are deviations from the recommendation in clinical practice.
	Staging test thresholds	**Optimal:** <120 mmHg systolic and/or <80 mmHg diastolic arterial pressure **Normal:** 120-129 mmHg systolic and/or 80-84 mmHg diastolic arterial pressure **High normal:** 130-139 mmHg systolic and/or 85-89 mmHg diastolic arterial pressure	**Neglectable practice variation:** No or some deviations from the recommendation are observed, however it was judged that these deviations are not of importance or that the observed deviations are not necessarily unwanted. **Considerable unwanted practice variation:** It was judged that most of the observed deviations from the recommendation are unwanted.
Management	*Option #1, when: indications*	**Do not provide an intervention, when:** *Optimal OR normal blood pressure is present*	**Re-assess at a later point in time, when:** *Neglectable practice variation is present*
	Option #2, when: indications	**Advice on lifestyle changes, when:** *High normal blood pressure is present*	**Archive, when:** *Neglectable practice variation is present AND the clinical field signals that guidance is no longer needed**
	Option #3, when: indications	**Advice on lifestyle changes and consider drug prescription, when:** *High normal blood pressure is present* AND *very high cardiovascular risk profile especially with coronary artery disease is present**	**Update, when:** *Considerable unwanted practice variation is present*

*Different staging variables and tests are required to provide enough information for the several indications to choose an appropriate management option.

in guideline development methodology handbooks (e.g. updating procedures). Available management options can have a unique set of subsequent actions. For instance, archiving a CPG requires different actions compared to withdrawing or (not) updating a CPG. Additionally, it's worth considering that the set of management actions may differ between organizations for the same management option (e.g. updating).

3.2.4 MONITORING

In clinical practice, patients are usually followed over time to assess whether the selected management succeeded, to identify disease recurrence, or to assess disease progression. Similarly, SRs or CPGs in the portfolio can be monitored through cyclical assessments (see Figure A2 in Additional file 1). The cyclical assessments start by pre-specifying a time interval on which these reassessments take place. This means that the expiration of the prespecified time interval triggers a new cycle of assessments in the maintenance strategy rather than indicating that the SRs or CPGs are outdated. The choice of appropriate time intervals is essential. Prespecified time intervals should be long enough to allow for the development of new cases, recurrences, or progression, but not so long to cause excessive harm when the target condition had already developed early in the interval. If time intervals are too short, frequent assessments are resource intensive relative to the benefits. Too long intervals might lead to harmful consequences due to delayed identification of evolving conditions or outdated conclusions and recommendations.

3.3 DESIGNING AND TAILORING A MAINTENANCE STRATEGY

Maintenance strategies within organizations can potentially be designed and tailored according to the needs and capabilities of the organization by using the concepts of a test-treatment pathway. Table A10 in Additional file 1 provides a blank process description table to design or tailor a maintenance strategy. Some detection and staging variables could provide more predictive information than others. The measurement of information on those variables may require more resources due to the nature of the tests or test protocols involved. If the organization is not capable or willing to spend such resources (e.g. budget, work force, time), a less resource intensive variable, test, or test protocol may be selected to obtain the information. However, this trade-off might result in a reduced predictive strength for the presence or absence of the target condition and management indications. Three examples of tailored maintenance strategies are provided in Additional file 1 (Tables A11-13 and Figures A3-5, respectively)

In these hypothetical scenarios, different choices were made between strategies leading to variations in how the target condition was defined, the selection of different detection and staging variables and tests, differences in management indications, and the availability of different management options. These variations resulted in different process flows, even though the underlying concepts and elements within the framework remain the same.

Table 6 – Examples of management options and their indications in the presence of the target condition.

	Clinical healthcare example	CPG maintenance scenario
Target condition	Arterial hypertension is present (≥140/90 mmHg)	The recommendation is outdated (There was new peer-reviewed scientific evidence published)
Objective	To choose an appropriate management option in the presence of the target condition using the information gathered from staging	To choose an appropriate management option in the presence of the target condition using the information gathered from staging
Management option #1, when: indications	**Advice on lifestyle changes, when:** *Grade 1 blood pressure* AND *low to moderate cardiovascular risk profile without cardiovascular disease, renal disease, or hypertension-mediated organ damage**	**Update the recommendation, when:** *New evidence is very likely cause changes in the strength of the recommendation* OR *new evidence is very likely cause changes in the direction of the recommendation* OR *new evidence indicates a new recommendation should be added* OR *new evidence indicates an existing recommendation should be removed**
Management option #2, when: indications	**Advice on lifestyle changes and drug prescription after 3-6 months, when:** *Grade 1 blood pressure* AND *low to moderate cardiovascular risk profile without cardiovascular disease, renal disease, or hypertension-mediated organ damage* AND *blood pressure was not controlled after 3-6 months of lifestyle interventions**	**Do not update the recommendation, but reassess at a later point in time, when:** *The new evidence is very unlikely or somewhat unlikely to cause changes in the strength or direction of the recommendation* OR *no new evidence was found* OR *a valid justification is provided to postpone an update**
Management option #3, when: indications	**Advice on lifestyle changes and immediate drug prescription, when:** *(Grade 1 blood pressure* AND *high to very high cardiovascular risk profile with cardiovascular disease, renal disease, or hypertension-mediated organ damage)** OR *grade 2 blood pressure* OR *grade 3 blood pressure*	**Withdraw the recommendation, when:** *The clinical field signals that guidance is no longer needed* OR *new evidence indicates that an intervention should be de-implemented* OR registry data shows that the recommendations were fully implemented*

*Different staging variables and tests are required to provide enough information for the several indications to choose an appropriate management option. A sphygmomanometer or ambulatory blood pressure monitor only provides information about the level of blood pressure. Similarly, a survey for the CPG recommendation that was solely intended to assess possible changes in the strength and direction does not provide information about de-implementation or whether new recommendations should be added.

4. DISCUSSION

4.1 THE FRAMEWORK IN CONTEXT

Initially, we observed a large variety of updating strategies being reported in the literature.[3, 9-12, 15, 21-24] These strategies may not directly be applicable or adopted by other organizations, as organizations probably must consider various factors related to their context, capabilities, needs, and available resources when designing or tailoring their maintenance strategy. Different choices for those considerations may result in different strategies being implemented. The POMBYTT framework introduces key components in maintenance strategies based

on a diagnostic test-treatment pathway. It provides theoretical guidance to designers, emphasizing the explicit consideration of key elements in the framework and thus operational aspects in the strategy. First, it prompts consideration about how the target condition (e.g. outdatedness) is defined, ensuring clarity in its definition. Next, it guides the determination of how the presence or absence of the target condition is assessed, including establishing the threshold for decision-making. Furthermore, the framework guides considerations for selecting appropriate management options based on indications, how to test for these indications and establishing staging thresholds. Additionally, it guides considerations about how monitoring processes can be performed. The components and elements may also be useful for stakeholders and end-users of SRs and CPGs. For instance, understanding the diagnostic and staging components can be helpful for clinicians and local protocol developers to informally screen the CPGs and SRs they consult. This might eventually result in stronger signals from the clinical field to organizations maintaining SRs and CPGs, indicating whether an SR or CPG is considered outdated for practice.

Some of the reported strategies lead to multiple management options.[9, 10, 15, 16, 25] Most of these options seem to focus on variations of (not) updating. For example, *"don't update"*, "*don't update yet*", "*to be updated*", or "*update now",*[16] and *"prepare update"*, *"update pending"*, *"no update planned",* or *"up to date".*[15] Other strategies lead to *"exclude"*, *"no update"*, *"exceptional update"*, and *"start regular update",*[9] or *"don't update"*, "*don't update yet*", and "*to be updated*".[16] This may reflect the different needs or preferences for management options within organizations. Through the POMBYTT framework it becomes prevalent that there might be more management options available in the strategy than (not) updating, even when the target condition is absent. For example, re-endorsing, archiving, or withdrawing. The theoretical framework reveals that the question '*when to update?*' is only one part of a maintenance strategy, which leads to the updating management option. The question '*how to manage?*' is probably a more encompassing question in the context of portfolio maintenance. Furthermore, the framework could potentially aid in adapting existing strategies to the needs and capabilities of an organization. The existing strategy could be mapped to the framework (e.g. by using the Table A10 in Additional file 1) and changes or additions to the strategy can be made in line with the organization's context, needs, capabilities, and/or resources.

It can be argued that the living SR or CPG is a competing or complementary concept to the POMBYTT framework. However, it is possible to map the elements of living SRs or CPGs to the theoretical POMBYTT framework. In the case of a living CPG recommendation, updates are made when new relevant evidence becomes available.[18] Based on this, we can deduce that the definition of the target condition could be *'outdatedness of a recommendation is present when there is new relevant evidence'*, the detection variable could be *'new evidence'*, the detection test could be a *'literature search and selection'*, and the detection threshold is *'any new relevant evidence'*. Further guidance suggests a possible staging test where the CPG panel discusses the potential effect of changes in the body of evidence on the recommendation.[26] This approach is also seen in other living CPG literature, where an expert panel could be considered as a staging test using '*the content of the recommendation changes* OR *the strength of the recommendation changes*' as management indications.[27] The guidance also provided management options for living CPG recommendations: no modification, modification of elements in the recommendation, merging recommendations,

splitting recommendations, retirement, and removal.[26] With Table A13 and Figure A5 (Additional file 1) we adapted information found in living recommendation literature[18, 26, 27] for illustrative purposes to provide a hypothetical example of a living strategy.

4.2 CONSIDERATIONS FOR VARIABLES

The needs and capabilities of an organization may be a factor in selecting detection and staging variables for a tailored maintenance strategy. However, literature may also provide some evidence about which variables to use. One study reported that both the '*number of new trials*' and the '*identification of new drugs*' were predictors for the decision to update SRs in a multivariable model.[28] The authors reported that '*a newly approved indication for an existing drug*' was not a significant predictor. Another study predicted the probability that conclusions would change in an update.[25] Three variables (i.e. *effect size ratio, I-squared, power*) were not significant predictors in univariable analyses. Six variables were significant predictors in univariable analyses while only the '*number of new trials*' and the '*log weight ratio*' remained in the multivariable model predicting changes in conclusions. The exclusion of the four other variables (i.e. *large new trial, log participant ratio, logit standard error, log study ratio*) in the multivariable analysis indicates that these variables carried less predictive information. Variables containing less predictive information might still be good enough as proxy variables when organizations are unable to spend their resources for obtaining data on the known best predictors.

4.3 CONSIDERATIONS FOR TESTS

Different tests and test protocols may provide information with different predictive strength on the same variable. Surveying experts for new evidence is arguably less resource intensive than performing a systematic literature search and selection. However, a systematic approach of search and selection might yield higher predictive information in terms of the number of identified studies. Systematic searches and selections might not be feasible for resource-limited organizations. Especially when individual searches and manual literature selections are performed for every key question in the organization's portfolio. This might change in the future when machine learning systems are deployed to reduce time investments.[29, 30] Nevertheless, the gained time investments from semi-automation currently might come at a loss of accuracy in the study selection.[31, 32]

Even within a single test there could be a difference in the resulting predictive information as variations could arise in the test protocol. For example, a single-person literature screening and selection might result in more missed studies than an independent double-blind literature screening. Other examples of variations within literature search and selection protocols in favor of time efficiency can be found in rapid review methodology, where it is proposed to dual screen at least 20% of the abstracts.[33] Future considerations about the impact on the predictive quality of information in test protocols might include whether single or dual-person literature selections are assisted by machine learning systems. Currently, semi-automating the literature selection in a single person protocol could result in a larger risk of missing relevant literature in the selection.[34]

4.4 CONSIDERATIONS FOR MONITORING

Conclusions and recommendations seem to get out of date at variable rates,[1, 2] thus a prespecified time interval itself does not inform which specific SR or CPG needs maintenance. The function of a prespecified time interval in the POMBYTTS framework, rather, is to initiate a new cycle of (re)assessments. Cyclical monitoring can enable the detection of new developments, recurrences, and progression. To detect a recurrence, the target condition needs to be present again after previous management actions were initially carried out to resolve the presence of the target condition. However, in some circumstances the target condition may be present in the SR or CPG but is not severe enough to allocate resources to for further maintenance actions, such as updating. Cyclical monitoring could then be used to monitor the progression of the target condition over time until the threshold is reached and indications for the management option are present. For example, when new evidence is available and does warrant new recommendations or a change the direction or strength of the recommendation. Here, the target condition can be present but no indications for updating are present. Future reassessments may show that the threshold is reached, indicating an update is appropriate. Setting an appropriate time interval between reassessments could be difficult. The interval should be long enough for the target condition to develop or progress but short enough to do no excessive harm when the target condition already developed or progressed early. A living CPG concerning pharmacological interventions for neuropathic pain after spinal cord injury searched for new evidence after 21 months and 10 months thereafter, respectively.[27] The living SR[35] in the World Health Organization's '*Therapeutics and COVID-19'* guideline[36] monitored the literature daily. The interval may be dependent on the rate of developments in the specific field, available resources, or urgency.

4.5 LIMITATIONS

One limitation of the presented framework is that it remains theoretical and has not yet been piloted in real-world situations for the development of SR and CPG maintenance strategies. While current updating and maintenance strategies can be mapped to the framework, its practical implementation and usability have not been tested. This is particularly relevant when dealing with very large portfolios, as monitoring the entire portfolio can be resource intensive. To address this challenge, one potential solution is to select less resource intensive tests that still provide an acceptable level of predictive information.

Another limitation pertains to the search and selection of the literature for our review. The search strategy primarily focused on identifying literature related to updating, and other maintenance options were not specifically targeted. Additionally, only literature that reported at least one indicator for the need for updating was included, potentially excluding literature solely reporting considerations for alternative management options. However, this limitation mainly affects the extent of examples provided and does not impact the fundamental concepts and elements of the framework.

Furthermore, subjective decisions were made during the selection of literature. For instance, some processes were categorized as need for updating processes rather than prioritization processes.[16, 21, 25, 37] The examples of variables, tests, and thresholds in Additional file 1 were

based on our interpretation for elements in the framework and may not align with the intended use in the original publications.

4.6 FUTURE DIRECTIONS

In the future, there is potential for an evidence ecosystem to emerge, connecting the primary research community, the evidence synthesis community, the guideline developing community, and their stakeholders.[38] Processes within organizations participating in the ecosystem need to assure that exchangeable products and cocreated products are trustworthy. In our opinion, this is two-fold: trustworthy in terms of quality (due to rigorous development procedures) and trustworthy in terms of up-to-date products (due to rigorous portfolio maintenance strategies). The current theoretical POMBYTT framework might be a valuable tool to potentially design or adapt maintenance strategies for organizations in an evidence ecosystem to keep their SRs or CPGs up-to-date. This might particularly be important for resource-constrained organizations who face challenges in allocating resources for maintenance activities. In an ideal world, using the maintenance framework results in a strategy where the whole portfolio can enter a maintenance strategy and receive appropriate management actions by selecting less resource intensive tests. However, organizations with limited resources could also use priority-setting assessments to spend the available resources for maintenance on those SRs or CPGs with the highest priority. This requires new concepts to be introduced to the current theoretical POMBYTT framework.

The two hypothetical strategies designed with the framework (Tables A11-12 and Figures A3-4 in Additional file 1) and the living strategy derived from information from living recommendation literature[18, 26, 27] mapped to the framework's elements (Table A13 and Figure A5 in Additional File 1) might demonstrate the framework's potential applicability and relevance for maintenance practices. However, the POMBYTT framework has not undergone empirical validation in real practice. Therefore, future research could focus on potential application in research and practice by assessing the usability and feasibility of the POMBYTT framework for designing maintenance strategies and thereafter assessing the feasibility of the designed strategy for maintaining a portfolio or SRs or CPGs in the real-world. Research within the scope of the framework could focus on identifying detection and staging variables with acceptable predictive qualities given the resources available to obtain data on these variables. Artificial intelligence might enable the use of sensitive literature search strategies while relieving the workload associated with literature selections. Organizations may then choose to reallocate freed up resources to improve other test protocols that could provide better predictive information but are more resource intensive.

5. CONCLUSIONS

The choices regarding variables, tests, test protocols, indications, management options, and monitoring when designing a maintenance strategy with the theoretical POMBYTT framework will have a direct impact on the resulting processes in the strategy. These

elements aid in thinking about and being explicit about how the strategy operates when designing a maintenance strategy. For the resource-constrained organization it seems important to consider what result in acceptable predictive information about the presence or absence of the target condition and management indications while minimizing the resource investments. Understanding the components in the framework may also be helpful for stakeholders and end-users of SRs and CPGs to informally screen whether the SR or CPG is potentially still valid. Although the theoretical POMBYTT framework needs testing in the real world, it highlights important elements that should be explicitly considered when designing or adapting maintenance strategies. By taking these elements into account, organizations might potentially develop maintenance strategies related to their needs and context. Furthermore, the framework shows that there can be multiple management options available within a strategy, even when the target condition is absent. This highlights the importance of considering alternative management options beyond solely focusing on updating, probably offering greater flexibility in maintenance approaches.

LIST OF ABBREVIATIONS

CPG: Clinical Practice Guideline
POMBYTT: Portfolio Maintenance by Test-Treatment
SR: Systematic Review

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

All data generated or analyzed during this study are included in this published article and its supplementary information file.

COMPETING INTERESTS

All authors have completed the ICMJE disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare:

M.S. Oerbekke, Msc: No conflicts of interest

R.G. Elbers, PhD: No conflicts of interest
M.J. Van der Laan, MSc, MD, PhD: No conflicts of interest
L. Hooft, PhD: No conflicts of interest

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

All author contributions are made transparent using the structured Contributor Role Taxonomy (CRediT) described at https://doi.org/10.1002/leap.1210.

MSO: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, Writing – review & editing.
RGE: Investigation, Writing – review & editing.
MJvdL: Conceptualization, Methodology, Writing – review & editing.
LH: Conceptualization, Methodology, Writing – review & editing.

ACKNOWLEDGEMENTS

We would like to thank Dr. M. Ghannad for her contributions to this work (validation, data curation).

LITERATURE LIST

1. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147(4):224-33.
2. Martínez García L, Sanabria AJ, García Alvarez E, Trujillo-Martín MM, Etxeandia-Ikobaltzeta I, Kotzeva A, et al. The validity of recommendations from clinical guidelines: a survival analysis. Cmaj. 2014;186(16):1211-9.
3. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? Bmj. 2001;323(7305):155-7.
4. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. Cochrane Database Syst Rev. 2008(1):Mr000023.
5. Cumpston M, Chandler J. Chapter IV: Updating a review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021): Cochrane; 2021.
6. Thomas J, Askie LM, Berlin JA, Elliott JH, Ghersi D, Simmonds M, et al. Chapter 22: Prospective approaches to accumulating evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021): Cochrane; 2021.
7. Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, Moher D. AHRQ Technical Reviews. Updating Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007.
8. Shekelle P, Newberry S, Maglione M, Shanman R, Johnsen B, Carter J, et al. AHRQ Methods for Effective Health Care. Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005–2009). Rockville (MD): Agency for Healthcare Research and Quality (US); 2009.
9. Becker M, Neugebauer EA, Eikermann M. Partial updating of clinical practice guidelines often makes more sense than full updating: a systematic review on methods and the development of an updating procedure. J Clin Epidemiol. 2014;67(1):33-45.
10. Becker M, Jaschinski T, Eikermann M, Mathes T, Bühn S, Koppert W, et al. A systematic decision-making process on the need for updating clinical practice guidelines proved to be feasible in a pilot study. J Clin Epidemiol. 2018;96:101-9.
11. Agency for Health Care Policy and Research. Process for determining need for updates of clinical practice guidelines. Fed Regist. 1994;59(79 Pt 1):19723-5.
12. Gartlehner G, West SL, Lohr KN, Kahwati L, Johnson JG, Harris RP, et al. Assessing the need to update prevention guidelines: a comparison of two methods. Int J Qual Health Care. 2004;16(5):399-406.
13. Martínez García L, Arévalo-Rodríguez I, Solà I, Haynes RB, Vandvik PO, Alonso-Coello P. Strategies for monitoring and updating clinical practice guidelines: a systematic review. Implement Sci. 2012;7:109.
14. Martínez García L, Pardo-Hernandez H, Superchi C, Niño de Guzman E, Ballesteros

M, Ibargoyen Roteta N, et al. Methodological systematic review identifies major limitations in prioritization processes for updating. J Clin Epidemiol. 2017;86:11-24.
15. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, et al. When and how to update systematic reviews: consensus and checklist. Bmj. 2016;354:i3507.
16. Welsh E, Stovold E, Karner C, Cates C. Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach. J Clin Epidemiol. 2015;68(3):341-6.
17. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30.
18. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol. 2017;91:47-53.
19. Gopalakrishna G, Langendam MW, Scholten RJ, Bossuyt PM, Leeflang MM. Defining the clinical pathway in cochrane diagnostic test accuracy reviews. BMC Med Res Methodol. 2016;16(1):153.
20. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. 2018;39(33):3021-104.
21. Agbassi C, Messersmith H, McNair S, Brouwers M. Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. J Clin Epidemiol. 2014;67(12):1335-42.
22. Ahmadzai N, Newberry SJ, Maglione MA, Tsertsvadze A, Ansari MT, Hempel S, et al. A surveillance system to assess the need for updating systematic reviews. Syst Rev. 2013;2:104.
23. Chung M, Newberry SJ, Ansari MT, Yu WW, Wu H, Lee J, et al. Two methods provide similar signals for the need to update systematic reviews. J Clin Epidemiol. 2012;65(6):660-8.
24. Pattanittum P, Laopaiboon M, Moher D, Lumbiganon P, Ngamjarus C. A comparison of statistical methods for identifying out-of-date systematic reviews. PLoS One. 2012;7(11):e48894.
25. Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. Bmj. 2013;347:f7191.
26. El Mikati IK, Khabsa J, Harb T, Khamis M, Agarwal A, Pardo-Hernandez H, et al. A Framework for the Development of Living Practice Guidelines in Health Care. Ann Intern Med. 2022;175(8):1154-60.
27. Bragge P, Allison DJ, Mehta S, Guy S, Loh E. A pilot living clinical practice guidelines approach was feasible and acceptable to guideline panel members. J Clin Epidemiol. 2022;143:22-9.
28. Peterson K, McDonagh MS, Fu R. Decisions to update comparative drug effectiveness reviews vary based on type of new evidence. J Clin Epidemiol. 2011;64(9):977-84.
29. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8(1):163.
30. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017;91:31-7.

31. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, Hartling L. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst Rev. 2019;8(1):278.
32. Reddy SM, Patel S, Weyrich M, Fenton J, Viswanathan M. Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence. Syst Rev. 2020;9(1):243.
33. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021;130:13-22.
34. Gates A, Gates M, Sebastianski M, Guitard S, Elliott SA, Hartling L. The semi-automation of title and abstract screening: a retrospective exploration of ways to leverage Abstrackr's relevance predictions in systematic and rapid reviews. BMC Med Res Methodol. 2020;20(1):139.
35. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. Bmj. 2020;370:m2980.
36. Lamontagne F, Agarwal A, Rochwerg B, Siemieniuk RA, Agoritsas T, Askie L, et al. A living WHO guideline on drugs for covid-19. Bmj. 2020;370:m3379.
37. Soll RF. Updating reviews: the experience of the Cochrane Neonatal Review Group. Paediatr Perinat Epidemiol. 2008;22 Suppl 1:29-32.
38. 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. J Clin Epidemiol. 2020;123:153-61.

CHAPTER 4
Additional files

ADDITIONAL FILE 1

A1.1 REVIEW METHODS

A formal literature review protocol was not prepared a priori and the literature review was not registered.

A1.1.1 SEARCH STRATEGY

We searched MEDLINE (via PubMed) for relevant published literature on 10 October 2018 which was updated on 29 April 2020. The search string contained keywords with synonyms and was constructed as follows: ((need for updating) OR (prioritizing)) AND (methods) AND (systematic review OR guideline). Additionally, PubMed Health (discontinued on 31 October 2018 [1]) was searched for relevant reports on 10 October 2018 with a similar search string. Hits from PubMed Health were limited via the search engine's interface by using the 'methods resources' filter. The complete search string for PubMed and PubMed Health are available under subheadings A1.1.1.1 and A1.1.1.2. We used a previously published SR to identify handbooks that considered updating CPGs [2]. If available, the latest version of a handbook was obtained through the organization's website. All reference lists from the included journal articles, reports, and handbooks were hand searched for additional relevant literature.

A1.1.1.1 PUBMED SEARCH STRING

((need[tiab] OR needs[tiab] OR needed[tiab] OR signal[tiab] OR signals[tiab] OR require[tiab] OR requires[tiab] OR required[tiab] OR detect[tiab] OR when[tiab]) AND (update[tiab] OR updates[tiab] OR updating[tiab] OR updated[tiab] OR "out of date"[tiab])) OR (prioritize[tiab] OR prioritise[tiab] OR prioritization[tiab] OR prioritisation[tiab] OR priority[tiab] OR priorities[tiab] OR prioritized[tiab] OR prioritised[tiab] OR prioritizing[tiab] OR prioritising[tiab]) AND ("Models, Theoretical"[Mesh] OR method[tiab] OR methods[tiab] OR methodology[tiab] OR methodologies[tiab] OR process[tiab] OR processes[tiab] OR surveillance[tiab] OR strategy[tiab] OR strategies[tiab] OR system[tiab] OR systems[tiab] OR approach[tiab] OR approaches[tiab] OR criteria[tiab] OR checklist[tiab] OR tool[tiab] OR content[tiab]) AND ("Review literature as topic"[mesh] OR "meta-analysis as topic"[mesh] OR "Practice guidelines as topic"[mesh] OR "guidelines as topic"[mesh] OR "Time factors"[mesh])

A1.1.1.2 PUBMED HEALTH SEARCH STRING

FILTER: METHODS RESOURCES

((need[tiab] OR needs[tiab] OR needed[tiab] OR signal[tiab] OR signals[tiab] OR require[tiab] OR requires[tiab] OR required[tiab] OR detect[tiab] OR when[tiab]) AND (update[tiab] OR updates[tiab] OR updating[tiab] OR updated[tiab] OR "out of date"[tiab])) OR (prioritize[tiab] OR prioritise[tiab] OR prioritization[tiab] OR prioritisation[tiab] OR priority[tiab] OR priorities[tiab] OR prioritized[tiab] OR prioritised[tiab] OR prioritizing[tiab] OR prioritising[tiab])

A1.2 ELIGIBILITY AND LITERATURE SELECTION

Studies, reports, and handbooks that described at least one signal or consideration to assess the need for updating of SRs or (sections of) CPGs were included. We did not consider time thresholds (e.g. “update after 3 years”) as signals that inform decisions in the assessment of the need for updating. We believed that time thresholds immediately trigger an update without providing information about the actual outdatedness and therefore do not inform decisions in assessing the need for updating. Literature was excluded when they were non-healthcare related, conference abstracts, oral or poster presentations, or when literature could not be obtained. Obvious non-relevant hits regarding unrelated topics were excluded by one author (MSO). When in slightest doubt the reference was advanced to the title and abstract screening phase. The resulting hits were screened on title and abstract by two authors (MSO, RGE) blinded for each other’s decision. The reference was advanced to the full text selection phase when in doubt or disagreement. Full text articles, reports, and handbooks were read and selected by two authors (MSO, RGE) while blinded for each other’s decision. When no consensus could be reached or unclarity persisted after reading the full text, a third author (LH) was consulted for a final decision.

A1.3 DATA EXTRACTION AND DATA HANDLING

One author (MSO) extracted the data using a standardized data-extraction form. General characteristics such as publication year, author, organization, country and title were extracted and if the indicator was described for an SR or CPG. Signals and considerations specifying an organization, field, or profession were replaced with generalized terms. For example, “New serious safety alert issued by the FDA or Health Canada” [3], was generalized to “New serious safety alert issued by federal organizations”. Multiple considerations or signals presented in one sentence or as a cluster were separated when assumed that they could be independent of each other. For example, “The nature and volume of new evidence” [4] was separated into “The nature of new evidence” and “The volume of new evidence”. Extracted signals and considerations were cross-checked by another researcher.

A1.4 DATA ANALYSIS

We used NVivo for Windows (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12) to identify common themes among the extracted data. One author (MSO) performed the qualitative coding. The emerging domains were discussed with one author (LH). Changes as a result of the discussion were made without the use of NVIVO. Additional data extracted from papers identified in the updated search strategy were placed under the existing domains. The extracted data was thereafter characterized as detection variables, detection tests, detection thresholds, staging variables, staging tests, staging thresholds, or management indications at our own discretion.

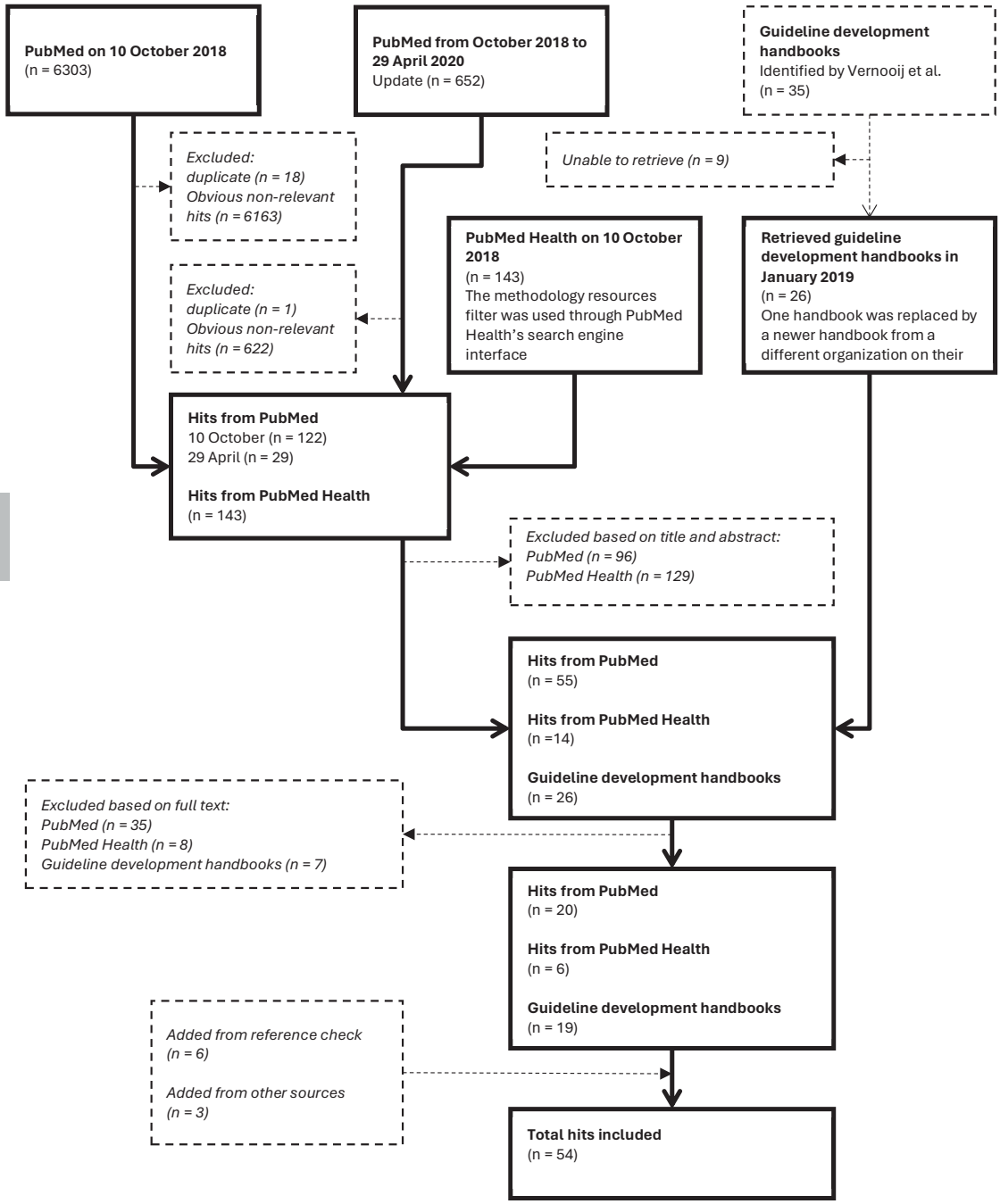


Figure A1. Flow diagram of the study selection. *Handbooks were identified through a previously published systematic review by Vernooij et al. [2]*

Table A1 – Reasons for exclusion

First author / organization	Year	Title	Reason for exclusion
Agencia d'avaluacio de technologia i recerca mediques		Guies de pràctica clínica.	Could not be found online
Akl	2017	The SPARK Tool to prioritize questions for systematic reviews in health policy and systems research: development and initial validation.	No need for updating indicators were found
Akl	2017	Living systematic reviews: 4. Living guideline recommendations	No need for updating indicators were found: refers to Chung 2012 and Shekelle 2014 for thresholds for when to update.
Alderson	2014	Median life span of a cohort of National Institute for Health and Care Excellence clinical guidelines was about 60 months	No need for updating indicators were found
Alonso-Coello	2011	The updating of clinical practice guidelines: insights from an international survey	No need for updating indicators described
American College of Chest Physicians		Evidence-based Guideline Development Process	No need for updating indicators were found
American College of Physicians	2010	The development of clinical practice guidelines and guidance statements of the American College of Physicians: summary of methods.	No need for updating indicators were found
American society of clinical oncology		American Society of Clinical Oncology Guideline Procedures Manual.	Could not be found online
Arzneimittelkommission der deutschen ärztschaft	2011	Leitfaden für die Erstellung von Therapieempfehlungen.	No need for updating indicators were found
Ärztliche Zentralstelle Qualitätssicherung		National Disease Management Guidelines: Method Report.	Could not be found online
Atkins	2012	Priority setting in guideline development: article 2 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report	Seems to cover new topics
Bastian	2011	Choosing health technology assessment and systematic review topics: the development of priority-setting criteria for patients' and consumers' interests	Seems to cover new topics
Battista	1995	Setting priorities and selecting topics for clinical practice guidelines	Seems to cover new topics/guidelines
Bero	2013	The Cochrane Collaboration review prioritization projects show that a variety of approaches successfully identify high-priority topics	No need for updating indicators were found

Bundersärtzekammer		National Disease Management Guidelines.	Could not be found online
Burgers	2012	Adaptation, evaluation, and updating of guidelines: article 14 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report	No need for updating indicators were found
Canadian thoracic society	2007	Canadian Thoracic Society: Presenting a new process for clinical practice guideline production.	No need for updating indicators were found
Centrul Naţional de Studii Medicina Familiei		Metodologie elaborarii ghidului de practica.	No need for updating indicators were found
Chalmers	1993	preparing and updating systematic reviews of randomized controlled trials of health care	No need for updating indicators were found
Clark	2006	From outdated to updated, keeping clinical guidelines valid	No original data
Dalal	2012	A Pilot Study Using Machine Learning and Domain Knowledge To Facilitate Comparative Effectiveness Review Updating [Internet]	No need for updating indicators were found
Domus Medical Flemish College of General Practitioners		Algemeen Stramien voor de Ontwikkeling van Aanbevelingen van Goede Medische Praktijkvoering.	Could not be found online
Doyle	2005	Global priority setting for Cochrane systematic reviews of health promotion and public health research	No need for updating indicators were found
Drug Commission of the German Medical Association	2006	Handbuch zur Entwicklung regionaler Leitlinien.	No need for updating indicators were found
Duodecim Finnish Medical Society		Submitted NHS Evidence Accreditation Application.	Could not be found online
Duodecim Medical Publications		Preface: What is Evidence-Based Medicine Guidelines.	Could not be found online
El-Harakeh	2019	Prioritization approaches in the development of health practice guidelines: a systematic review	No need for updating indicators were found
European Region Of The World Confederation For Physical Therapy		Framework for Clinical Guideline Development in Physiotherapy.	Could not be found online
French	2005	Investing in updating: how do conclusions change when Cochrane systematic reviews are updated?	No need for updating indicators were found
Gartlehner	2004	Assessing the need to update prevention guidelines: a comparison of two methods	No need for updating indicators were found

Handoll	2013	A framework for effective collaboration between specialist and broad-spectrum groups for delivering priority Cochrane reviews	Seems to cover new topics/No need for updating indicators were found
Hoomans	2012	Systematizing the Use of Value of Information Analysis in Prioritizing Systematic Reviews [Internet]	No need for updating indicators were found
Jiang	2016	Essential methods and procedures on how to develop and update clinical practice guidelines.	Unobtainable, doi: 10.3760/cma.j.issn.0376-2491.2016.04.004
Jiang	2019	Guideline for [Clinical Guidelines Constitution/Amendment] in China	No need for updating indicators were found
Joanna Briggs Institute Synthesis Science Unit		Best Practice Information Sheet (BPIS) Procedures.	Could not be found online
Jones	2013	Success for a novel approach to priority setting in South Australian public dental clinics	Seems to cover health care priority setting
Kim	2018	Identifying and prioritizing topics for evidence-based geriatric nursing practice guidelines in Korea	No need for updating indicators were found
Land	2017	A five-step approach for stakeholder engagement in prioritisation and planning of environmental evidence syntheses	Not healthcare related
Manafo	2018	Patient and public engagement in priority setting: A systematic rapid review of the literature.	Priority setting in health research and health eco-system
Martinez Garcia	2012	Strategies for monitoring and updating clinical practice guidelines: a systematic review	No original data
Martinez Garcia	2014	Updated recommendations: an assessment of NICE clinical guidelines	No need for updating indicators were found
Martinez Garcia	2017	Methodological systematic review identifies major limitations in prioritization processes for updating	No original data
McClarey	1999	Identifying priorities for national clinical guidelines	Seems to cover new topics
Moher	2008	When and how to update systematic reviews	Duplicate (PubMed) / No original data
Moher	2007	A systematic review identified few methods and strategies describing when and how to update systematic reviews	No original data
Moher	2008	When and how to update systematic reviews	No original data
Nasser	2013	An equity lens can ensure an equity-oriented approach to agenda setting and priority setting of Cochrane Reviews	No need for updating indicators were found

Nasser	2013	Ensuring relevance for Cochrane reviews: evaluating processes and methods for prioritizing topics for Cochrane reviews	No need for updating indicators were found
Nast	2019	Prioritizing topics in guideline development: results of a two-phase online survey of dermatologist members of the EADV	No need for updating indicators were found
Neuman	2014	Durability of class I American College of Cardiology/American Heart Association clinical practice guideline recommendations	No need for updating indicators were found
New Zealand Guidelines Group		Handbook for the Preparation of Explicit Evidence-base Clinical Practice Guidelines.	Could not be found online
NICE	2013	Interim Process and Methods Guide for the Clinical Guideline Updates Using Standing Committees Pilot Programme 2013 [Internet]	No need for updating indicators were found
NICE	2013	Interim Process and Methods of the Highly Specialised Technologies Programme [Internet]	No need for updating indicators were found
NICE	2014	Interim Methods Guide for Developing Service Guidance 2014 [Internet]	No need for updating indicators were found
Olav Vandvik	2014	[A new generation of reliable clinical practice guidelines through MAGIC]	No need for updating indicators were found
Pieper	2019	[Increasing the efficiency of guideline production: a narrative review]	No original data
Reveiz	2010	Prioritization strategies in clinical practice guidelines development: a pilot study	No need for updating indicators were found
Robinson	2015	Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies	No original data
Sampson	2008	Surveillance search techniques identified the need to update systematic reviews	No need for updating indicators were found
Sanabria	2020	Prioritizing clinical guideline questions for updating: the UpPriority Tool	No need for updating indicators were found
Scott	2018	Cochrane acute respiratory infections group's stakeholder engagement project identified systematic review priority areas	No need for updating indicators were found
Sibbald	2009	Priority setting: what constitutes success? A conceptual framework for successful priority setting	No need for updating indicators were found
Stead	2001	Updating a systematic review – what difference did it make? Case study of nicotine replacement therapy	No need for updating indicators were found

Synnot	2019	Selecting, refining and identifying priority Cochrane Reviews in health communication and participation in partnership with consumers and other stakeholders	No need for updating indicators were found
Taylor	2019	Use of multi-attribute decision-making to inform prioritization of Cochrane review topics relevant to rehabilitation	No need for updating indicators were found
Therapeutic Guidelines Limited	2017	How Therapeutic Guidelines are produced.	No need for updating indicators were found
Tsertsvadze	2011	Updating Comparative Effectiveness Reviews: Current Efforts in AHRQ's Effective Health Care Program	Duplicate (PubMed) / No original data
Tsertsvadze	2011	Updating comparative effectiveness reviews: current efforts in AHRQ's Effective Health Care Program	No original data
Tugwell	2013	Methods for setting priorities in systematic reviews	No need for updating indicators were found
Vernooij	2014	Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks	No original data
Voisin	2008	Strategies in assessing the need for updating evidence-based guidelines for six clinical topics: an exploration of two search methodologies	Concerned search methodologies
Wale	2013	The Cochrane Library review titles that are important to users of health care, a Cochrane Consumer Network project	No need for updating indicators were found
Zarnke	2000	A novel process for updating recommendations for managing hypertension: rationale and methods	Unobtainable, could not be found on the journal's website

Table A2 – General characteristics of included studies

Author or organization[†]	Year	Country[‡]	Brief description of the need for updating process
Agbassi et al. [5]	2014	Canada	A questionnaire with criteria categorizes guidelines in four categories: endorse, defer, review, archive. Guidelines are assessed by a methodologist and a clinical expert using the questionnaire. Guidelines flagged as 'review' were assessed by another questionnaire, which resulted in a decision: endorse, update, or archive.
Agency for Health Care Policy and Research [6]	1994	USA	A public meeting to address the need for updating and the timing of an update is held as soon as sufficient data is obtained that indicates an update may be needed.
Ahmadzai et al. [7]	2013	Canada	After a focused search, qualitative and quantitative signals were assessed. Experts could indicate if the conclusions were still valid and could provide any new references that might invalidate conclusions. Experts could also provide any references that were important to the topic, but did not invalidate conclusions. Safety alerts were assessed. Evidence was combined with expert opinion and conclusions were categorized: up-to-date, possibly out-of-date, probably out-of-date, or out-of-date.
American Academy of Orthopaedic Surgeons [8]	2011	USA	Guidelines that are at least five years old are listed. A committee makes the decision whether to update or not based on criteria.
American College of Cardiology Foundation and American Heart Association [9]	2010	USA	A research analyst and the chair compare current recommendations against the latest data. The writing committee is surveyed whether (parts of) a guideline needs updating. Late-breaking trials were reviewed and regulatory bodies were monitored as well. The necessity of a guideline review is then determined.
American College of Occupational and Environmental Medicine [10]	2017	USA	Literature reviews were performed periodically to identify major changes in the literature. Major changes would require more frequent focused updates.
American urological association [11]	2015	USA	A systematic review of literature published since the release of the guideline is performed. A panel determines whether a limited or full revision is warranted.
Barrowman et al. [12]	2003	Canada	A 'diagnostic' test was performed for whether meta-analyses were out of date based on the cut-off of the new participant ratio being greater than one.
Bashir et al. [13]	2018	Australia	Signals were assessed retrospectively in a cohort of updated systematic reviews. Literature included in the updated systematic review and trial registries were assessed to obtain updating signals.
Becker et al. [14]	2014	Germany	A systematic monitoring of red flags, information and commentaries by the guideline authors, guideline users or experts, and relevant evidence was described. Additional searches could be performed when necessary, followed by a final decision whether an update was indicated.

Becker et al. [15]	2018	Germany	A limited search was performed and new potentially relevant evidence was reported. The report was sent to the guideline group and an online survey was conducted. Results were summarized and a consensus meeting was held to determine which sections were in high need for updating. A multidisciplinary working group answered a set of questions in a survey. The survey results were analyzed and debated to determine which guidelines needed updating.
Chung et al. [16]	2012	USA	Two methods were used. The modified Ottawa-method used a focused literature search. New evidence was screened for their relevance to the review conclusions and qualitative, quantitative and 'other' indicators were assessed. The RAND-method used a limited literature search. Field experts were asked whether conclusions were still valid and to provide additional evidence.
Cumpston et al. [17]	2019	UK	A set of questions were answered sequentially. If any question was answered negatively, no update would be performed. Otherwise, an update would be performed.
Davis et al. [18]	2007	Canada	Persons familiar with the topic may conduct limited literature searches on a routine basis. The working group may identify literature that ensures a revision of the guideline.
Dumonceau et al. [19]	2012	Switzerland	NR§
Garner et al. [20]	2016	UK	A set of questions were answered sequentially. If any question was answered negatively, no update would be performed. Otherwise, an update would be performed.
Garritty et al. [21]	2010	Canada	NF§
Guidelines and Protocols Advisory Committee [22]	2017	Canada	NF§
Haller et al. [23]	2015	Belgium	NF§
Haute Autorité de Santé [24]	2016	France	NF§
Howell [25]	2015	Australia	A literature review from the date of the last search for the guideline is performed. Guideline writers are asked to appraise additional relevant studies and to determine whether recommendations may change (add or remove a recommendation, strength of a recommendation, and/or change a recommendation).
Iorio et al. [26]	2009	Italy	A coordinator may invite to draft an update when a relevant paper is published.
Javaher [27]	2015	USA	Epidemiologists review and revise. All potential changes and updates were presented to an advisory commission for their recommendations.
Johnston et al. [28]	2003	Canada	New evidence was searched. A working group was informed about the new evidence through a summary report and made a decision whether to update.
Kwaliteitsinstituut voor de Gezondheidszorg CBO [29]	2007	The Netherlands	Associations relevant to the clinical guideline are notified when data from several criteria shows that updates are important. Thereafter, it is decided whether an update is warranted.

Lyratzopoulos et al. [4]	2012	UK	Clinical experts could provide comments on whether there had been substantial changes in the evidence since the guideline was published. An updated search could be performed. The expert comments and the information retrieved by the literature search were discussed and a decision about updating the guideline was made.
Martínez García et al. [30]	2017	Spain	Each clinical question of the guideline was classified in one of the three categories: to be reviewed, valid, or new clinical question.
Martínez García et al. [31]	2014	Spain	New evidence that could change recommendations was gathered through expert identification and experts evaluated whether recommendations were still up to date. Furthermore, a literature search was performed. A decision for updating (i.e. valid or in need for updating) was made on references that could potentially trigger an update.
Meerhoff et al. [32]	2016	The Netherlands	Implementation problems are continuously monitored. Scientific expert from the previous working group inform about new literature since the guideline was published. The experts also keep informed about whether other organizations are planning to develop or update potentially relevant guidelines. Guideline experts advised the organization's board about the need to update one or more guidelines.
Mickenautsch et al. [33]	2013	South Africa	A modified Ottawa-method was used where a basic literature search was conducted. The qualitative, quantitative, and 'other' signals were assessed and assigned after study selection.
Murad et al. [34]	2015	USA	Periodically monitoring the literature for new evidence.
National Health and Medical Research Council [35]	2009	Australia	A multidisciplinary group (similar to the guideline development group) should assess the guidelines whether new evidence should be incorporated.
National Institute for Health and Care Excellence [36]	2014	UK	Without the need for a formal check, the need to update recommendations was assessed on a case-by-case basis when there are safety concerns. Decisions to update a guideline are based on the cumulative assessment of the evidence published since the guideline publication. Every two years the guidelines are assessed. Less resource-intensive checks are performed at 2/6/10-year timepoints, while extensive checks are performed at 4/8-year timepoints. Findings of the checks are discussed.
National Institute for Health and Clinical Excellence [37]	2013	UK	New evidence, views of the guideline development group and additional information about the relevance of the guideline is assessed. A decision was made on the surveillance review proposal: substantial update, rapid update, no update, transfer to static list, withdraw.
Newberry et al. [38]	2013	USA	Two methods were used separately or combined. The Ottawa-method had a focused literature search. New evidence was screened for relevance for the review conclusions. Qualitative and quantitative indicators were assessed and assigned. The RAND-method had a limited search. Experts responded to a questionnaire and could indicate whether conclusions were still valid. Experts also could provide new evidence. The evidence from the literature search and the expert opinion were then assessed.
Pattanittum et al. [39]	2012	Thailand	Several methods were compared. All methods were of statistical nature (or used quantitative indicators) and had predefined cut-off values. Methods included the assessment of sufficiency and stability, changes in effect size, the new participant ratio, changes in statistical significance, and simulation-based power.

Peterson et al. [3]	2011	USA	Based on a majority vote informed by new evidence and any variety of factors in addition to the evidence the decisions to update were made.
Rosenfeld et al. [40]	2013	USA	Reviewers were asked to complete a summary grid and to review each key statement. Thereafter, reviewers suggested: keep as is, keep but modify, or discard. Experts could also identify new evidence or quality improvement opportunities. Additionally, a literature search should be performed.
Scottish Intercollegiate Guidelines Network [41]	2015	UK	Individuals that comment on guidelines were asked to develop a small change proposal. The advisory group assessed small proposals and small changes may be agreed upon. The advisory group will also consider whether new evidence warrants a selected update or full review. A full proposal is then assessed together with new topics.
Shekelle et al. [42]	2014	USA	Literature was searched in databases and top-rated medical journals (general and topic specific specialty journals). Experts were asked to comment on whether the conclusion were out of date. Both the new evidence and expert opinions were assessed to determine whether it indicated a need for updating.
Shekelle et al. [43]	2001	USA	Experts were asked if there was new evidence and whether the new evidence was sufficient for an update. Experts were also asked whether there were new guideline statements (within the original scope) to be added. After that, a limited literature search was performed. Literature and expert opinion were reviewed for a final decision about the magnitude of an update: major update, minor update, or still valid.
Shekelle et al. [44]	2009	USA	A limited literature search was performed. Experts responded to a questionnaire and could indicate whether conclusions were still valid. Experts also could provide new evidence. The evidence from the literature search and the expert opinion were then assessed.
Shekelle et al. [45]	2014	USA	A limited search was performed. Experts responded to a questionnaire and could indicate whether conclusions were still valid. Experts also could provide new evidence. The evidence from the literature search and the expert opinion were then assessed.
Shekelle et al. [46]	2011	USA	Two methods were used. The Ottawa-method had a focused literature search. New evidence was screened for relevance for the review conclusions. Qualitative and quantitative indicators were assessed and assigned. The RAND-method had a limited search. Experts responded to a questionnaire and could indicate whether conclusions were still valid. Experts also could provide new evidence. The evidence from the literature search and the expert opinion were then assessed.
Shekelle et al. [47]	2001	USA	A limited search is proposed. Guideline recommendations are assessed by experts. Experts may indicate whether new recommendations should be present and whether they are aware of new evidence. An expert panel decides whether the guideline recommendation needs updating.
Shojania et al. [48]	2007	Canada	Literature was searched for systematic reviews and trials. Qualitative and quantitative indicators were assigned. Qualitative signals had two levels of importance: potentially invalidating, or major changes. A discussion was held for additional searches and the updating status was finalized.
Shojania et al. [49]	2007	Canada	A search for new systematic reviews and trials on the specific topic was performed. Candidate trials were screened and eligible studies were added to the existing meta-analysis. Qualitative and quantitative signals were assessed. An additional search could be performed when needed and the updating status was finalized.
Soll et al. [50]	2008	USA	Editors and regional coordinators assessed the need for updating based on criteria.

Sutton et al. [51]	2009	UK	One method used the new participants ratio, which indicates the number of new participants in the (null) meta-analysis to obtain significance. A second method was an adjusted power calculation. Here, simulations were used to calculate the proportion of simulations where the null hypothesis was rejected. Other indicators can be used in the simulation-based framework as well.
Takwoingi et al. [52]	2013	UK	A multi-component tool was used. First it was assessed whether the clinical question was already answered by the available evidence and whether it was still relevant. Then it was considered whether there were any new factors relevant to the review and it was assessed whether there were new studies. Availability of a review team was taken into account. Following a flow-chart, the decision could be made: to update now, to be updated, or to not update just yet.
US Preventive Task Force [53]	2015	USA	Based on indicators, a brief background paper was written. The paper was discussed and each topic was labeled as 'active' or 'inactive'.
Welsh et al. [54]	2015	UK	First it was assessed whether the clinical question already was still relevant. Then it was considered whether there were any new factors relevant to the review. It was also assessed whether there were new studies and whether they were likely to impact the review conclusions. Availability of a review team was taken into account. Following a flow-chart, the decision could be made: to update now, to be updated, or to not update just yet.
Working Group for CPG Updates [55]	2010	Spain	A limited literature search was performed. Guideline developers or experts were asked whether there was new relevant literature or new recommendations should be added. Furthermore, user needs and guideline context were assessed. A decision about whether or not to update was made based on the provided data.
World Health Organization [56]	2014	Switzerland	A continual assessment of how new information may affect recommendations in a rapid advice guideline. For recommendations that may be out of date in guidelines (based on new evidence), implementers and stakeholders are made aware of the uncertainty and plans for updating the recommendations.

[†]The first author or organization that was described was extracted.
[‡]The country mentioned in the first affiliation of the first author, or the country of the organization was extracted
[§]Not found or could not be deduced.
RAND: Research ANd Development, from the RAND Corporation
UK: United Kingdom
USA: United States of America

Table A3 – Examples from the literature review which could be used as detection variables

Domain	Item	Reference	Described for
New evidence	Emerging scientific evidence	[27]	Guideline
	Publication in a peer reviewed journal	[9]	Guideline
	Significant new clinical trials and/or peer reviewed literature on the topic	[9]	Guideline
	Nonrandomized data deemed important on the basis of results impacting safety and efficacy assumptions	[9]	Guideline
	Relevant paper published in the intervening period	[26]	Guideline
	Are participants of the multidisciplinary guideline working group aware of new potentially relevant evidence that was not identified by the limited search?	[15]	Guideline
	New evidence	[34]	Guideline
	Whether there is any new evidence that should be incorporated	[35]	Guideline
	New relevant evidence since the guideline publication	[36]	Guideline
	Abstracts of primary or secondary evidence that has been published since the end of the search period for the guideline, with critical appraisal of key papers	[36]	Guideline
	Information on important new evidence in the field	[41]	Guideline
	Has new evidence emerged since the original clinical practice guideline was drafted?	[55]	Guideline
	The updating of recommendations should be considered in light of data published in the scientific literature or significant changes in practice since the publication of the recommendations [Translated from French: L'actualisation des recommandations doit être envisagée en fonction des données publiées dans la littérature scientifique ou des modifications de pratique significatives survenues depuis la publication des recommandations]	[24]	Guideline
	Comments from the public regarding new scientific evidence or new technologies that may warrant the updating of a clinical practice guideline	[6]	Guideline
	Changes in the evidence that current practice is optimal	[40, 43, 47]	Guideline
	Changes in evidence on existing benefits and harms of interventions	[18, 22, 40, 47]	Guideline
	Changes in the evidence on the benefits and harms of existing interventions	[43]	Guideline
	In case of new evidence	[23]	

	The publication of a study that was included in the update of primary outcome meta-analysis within a year of the systematic review being published†	[13]	Systematic review
	The completion date of a study that was included in the update of primary outcome meta-analysis within a year of the systematic review†	[13]	Systematic review
	New information that is now available	[50]	Systematic review
	Missing information that is now available	[50]	Systematic review
	Information from existing studies (e.g. information about new treatment regimens, population subgroups, harms, economic data, or outcome measures, including data from ongoing studies or previously missing data)	[52]	Systematic review
	Additional information from existing studies	[54]	Systematic review
	Potentially relevant evidence from (limited) literature searches (if necessary, additional searches could be conducted)	[14]	Systematic review
	Are there any new studies, or new information?	[20]	Systematic review
	Reporting of serious or 'new' serious adverse events	[21]	Systematic review
	Identification of newly approved drug	[3]	Systematic review
	Identification of newly approved indication for previously included drug	[3]	Systematic review
Changes to the procedures, methods, resources, or contents	Changes in the available interventions	[18, 22, 40, 43, 47]	Guideline
	Changes in the outcomes that are considered important	[18, 22, 40, 43, 47]	Guideline
	Changes in the values placed on outcomes	[40, 43, 47]	Guideline
	Expanding or narrowing of the scope of the guideline [Translated from Dutch: Uitbreiding of inkrimping van de afbakening van de richtlijn]	[29]	Guideline
	Should the structure or the scope of the clinical practice guideline alter?	[15]	Guideline
	Are there new relevant subject areas which are not considered to date?	[15]	Guideline
	Do the questions and search criteria as they are in the document address current needs, such that an updated literature search would be useful and identify relevant evidence?	[5]	Guideline
	changes in health technologies	[27]	Guideline
	changes in procedures	[27]	Guideline
	Changes in the current practice [Translated from Dutch: Veranderingen in de huidige praktijk]	[29]	Guideline

	Changes in the resources available for health care	[18, 22, 40, 43, 47]	Guideline
	Has there been a sudden increase in costs and utilization?	[27]	Guideline
	New methodology (e.g. new statistical techniques, or changes is methodological guidance)	[52]	Systematic review
	New changes in methodology	[54]	Systematic review
	Are there any new relevant methods?	[20]	Systematic review
	New inclusion criteria (outcomes; interventions; populations, methodological advances/new analysis)	[21]	Systematic review
Alerts, feedback, requests, and comments	Any other pertinent factors provided by the clinical experts	[4]	Guideline
	Drugs and medical devices alerts	[30]	Guideline
	Changes/announcements/policies on both existing and emerging areas of disease assessment and treatment	[9]	Guideline
	Comments from healthcare providers on the guideline [Translated from Dutch: Commentaar van zorgverleners op de richtlijn]	[29]	Guideline
	All comments received on the organization's published guidelines	[41]	Guideline
	Request and requirements for review and update from the practice community, key stakeholders, and other sources of free relationships with industry or other potential bias	[9]	Guideline
	New serious safety alert issued by federal organizations	[3]	Systematic review
	Information and commentaries by the clinical practice guideline authors, other experts, and guideline users	[14]	Systematic review
	Alerts and information on medical product safety published by national authorities for medicine and current medical news published in medical newsletters	[14]	Systematic review
	Formal request from a policy or healthcare decision maker	[21]	Systematic review
	Response to user feedback	[52, 54]	Systematic review
Other	Previous recommendation statement	[53]	Guideline
	Recommendations of other guideline developers	[53]	Guideline

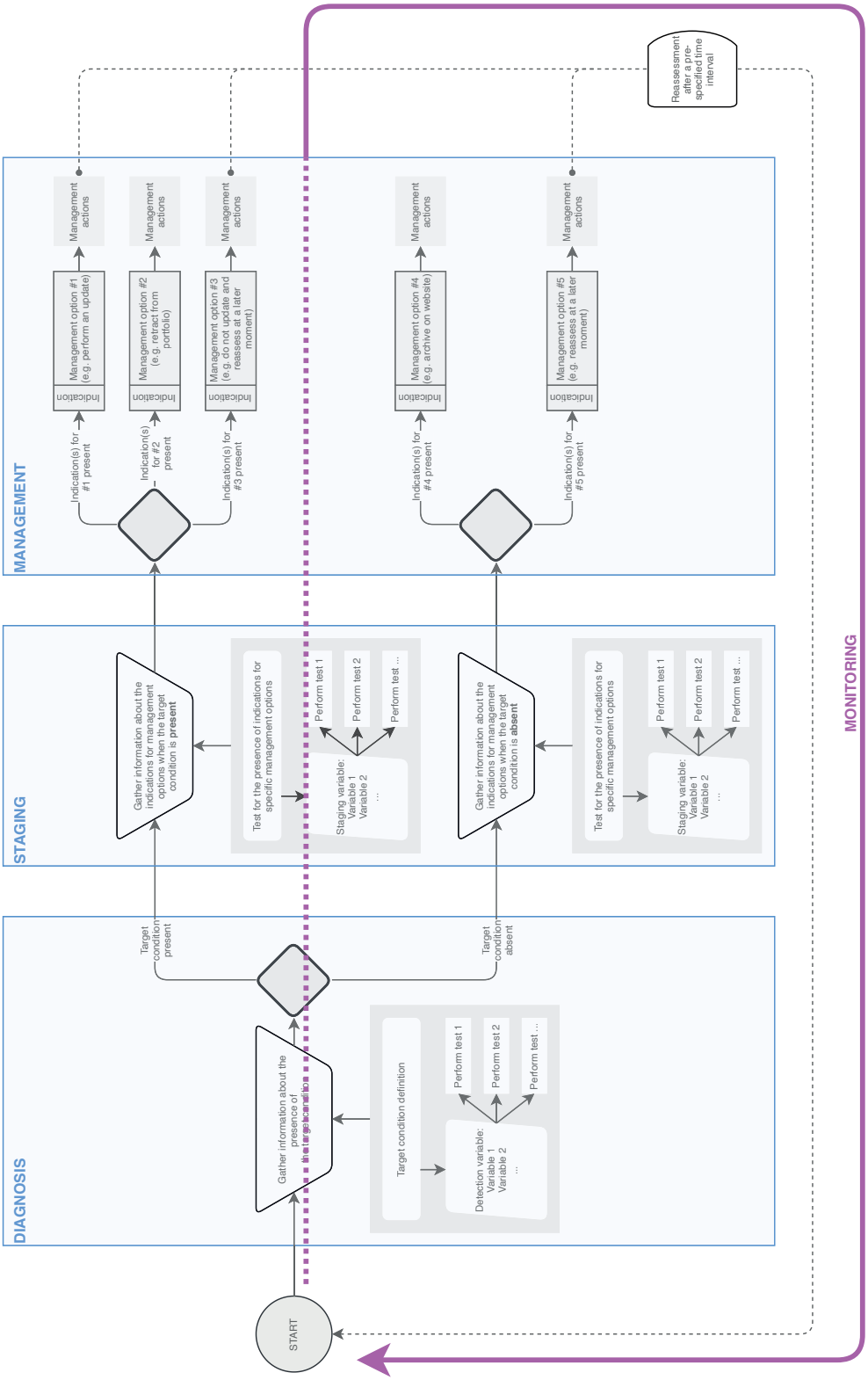

DIAGNOSIS
START
Gather information about the presence of the target condition
Target condition definition
Detection variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...
Target condition present
Target condition absent
STAGING
Gather information about the indications for management options when the target condition is present
Gather information about the indications for management options when the target condition is absent
Test for the presence of indications for specific management options
Staging variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...
MANAGEMENT
Indication(s) for #1 present
Indication(s) for #2 present
Indication(s) for #3 present
Indication(s) for #4 present
Indication(s) for #5 present
Indication
Management option #1 (e.g. perform an update)
Management option #2 (e.g. retract from portfolio)
Management option #3 (e.g. do not update and reassess at a later moment)
Management option #4 (e.g. archive on website)
Management option #5 (e.g. reassess at a later moment)
Management actions
Reassessment after a pre-specified time interval
MONITORING

Figure A2. The Portfolio Maintenance by Test-Treatment framework with outlined test-treatment concepts: *diagnosis, staging, management, and monitoring (outlined in blue). The purple line schematically represents the cyclical nature of monitoring. The dotted part of the purple line represents the (re)assessments in the maintenance strategy after a prespecified time interval.*

Table A4 – Examples from the literature review which could be used as detection tests

Domain	item	Reference	Described for
Expert and user solicitation	Guidelines staff periodically assess whether an existing guideline remains current through the update literature review process	[11]	Guideline
	Summary of brief literature search for new evidence	[53]	Guideline
	Intelligence gathering on the perceived current relevance of the guideline, which may include responses to questionnaires, information on guideline and quality standard implementation, external enquiries about the guideline recommendations, internal intelligence (such as an organization's guideline issues log), related guidance and quality standards (including placeholder statements in an organization's quality standards), medicines licensing information, relevant national policy	[36]	Guideline
	User perceptions: are any of the strategies or elements of the clinical practice guideline recommendations invalid?	[55]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: are you aware of new evidence relevant to the clinical practice guideline recommendations?	[55]	Guideline
	Context analysis: have there been any changes in the healthcare context of the clinical practice guideline?	[55]	Guideline
	Context analysis: are there any strategies or elements of the clinical practice guideline recommendations invalid?	[55]	Guideline
Traits of evidence	m/n as the "new participant ratio"	[12]	Systematic review
	The calculated power minus the proportion of significant results in the 10.000 iterations of re-meta-analyses	[39]	Systematic review
	The participant ratio (q) was calculated from q =m/n where m is the observed number of participants in the study(ies) published within the most recent 3 years, and n is the expected number of participants in the study(ies) published within the most recent 3 years	[39]	Systematic review
	Barrowman's n new participant ratio	[52]	Systematic review
	Participant ratio where the total number of participants (i.e. the total number in both the new and old studies) is compared to the total number in the old out-of-date meta-analysis	[52]	Systematic review
	Large new study where the total number of participants is greater than the total number in any of the studies in the out-of-date meta-analysis	[52]	Systematic review
	New pivotal study where the sample size is n-times that of any of the previous studies	[52]	Systematic review
	Standard error ratio where the standard error of the new effect size from the updated meta-analysis is compared to the standard error from the out-of-date meta-analysis	[52]	Systematic review

	Weight ratio where the total weight of the new studies is compared to the total weight of the old studies in the updated meta-analysis	[52]	Systematic review
	Effect size ratio whether the effect size in the updated meta-analysis is compared to the effect size in the out-of-date meta-analysis	[52]	Systematic review

Table A5 – Examples from the literature review which could be used as detection test thresholds

Domain	Item	Reference	Described for
New evidence	Recommendations were classified as still valid without key references. Key references were those that could potentially trigger an update.	[31]	Guideline
	If new evidence is published (For recommendations that may be out of date)	[56]	Guideline
	No new evidence or only confirmatory evidence and all responding experts assessed the conclusion as still valid [still up-to-date]	[7, 16, 38, 42, 44-46]	Systematic review
Traits of evidence	m/n>1 means a meta-analysis is out of date (for test: m/n as the "new participant ratio")	[12]	Systematic review
	Had to be ≥ 80% (for test: the calculated power minus the proportion of significant results in the 10.000 iterations of re-meta-analyses)	[39]	Systematic review
	q>1 (for test: The participant ratio (q) was calculated from q =m/n where m is the observed number of participants in the study(ies) published within the most recent 3 years, and n is the expected number of participants in the study(ies) published within the most recent 3 years)	[39]	Systematic review
	Ratio ≥ 5 (for test: Barrowman's n new participant ratio)	[52]	Systematic review
	Ratio ≥ 1.5 (for test: participant ratio where the total number of participants (i.e. the total number in both the new and old studies) is compared to the total number in the old out-of-date meta-analysis)	[52]	Systematic review
	Yes / no (for test: large new study where the total number of participants is greater than the total number in any of the studies in the out-of-date meta-analysis)	[52]	Systematic review
	n ≥ 3 (for test: new pivotal study where the sample size is n-times that of any of the previous studies)	[52]	Systematic review
	Ratio ≤ 0.5 (for test: standard error ratio where the standard error of the new effect size from the updated meta-analysis is compared to the standard error from the out-of-date meta-analysis)	[52]	Systematic review
	Ratio ≥ 1.5 (for test: Weight ratio where the total weight of the new studies is compared to the total weight of the old studies in the updated meta-analysis)	[52]	Systematic review

	Absolute slope of the linear regression >0 (for stability)	[39]	Systematic review
	Failsafe ratio >1 (for sufficiency)	[39]	Systematic review
	Ratio ≥ 1.5, ratio ≤ 0.5, i.e. change of 50% (for test: effect size ratio whether the effect size in the updated meta-analysis is compared to the effect size in the out-of-date meta-analysis)	[52]	Systematic review

Table A6 – Examples from the literature review which could be used as staging variables

Domain	item	Reference	Described for
Relevance	Is the document still relevant (clinically or to the care system as a whole in some way)?	[5]	Guideline
	Relevance to prevention and primary care	[53]	Guideline
	Did this guideline topic clarify practice or resolve an area of controversy?	[8]	Guideline
	Estimate of disease burden	[53]	Guideline
	Estimation of the clinical relevance	[15]	Guideline
	Does the published review still address a current question?	[20]	Systematic review
	The continuing importance of the review to decision makers	[17]	Systematic review
Traits of evidence	Strengths/weakness of research methodology and findings	[9]	Guideline
	The likelihood that such information (i.e. the volume and quality of new evidence) would cause a change in the guideline's recommendation	[6]	Guideline
	The availability of potentially relevant evidence that may influence the recommendations	[15]	Guideline
	Do the current recommendations cover all relevant subjects addressed by the evidence, such that no new recommendations are necessary?	[5]	Guideline
	Does newly identified evidence, on initial review, contradict the current recommendations, such that the current recommendations may cause harm or lead to unnecessary or improper treatment if followed?	[5]	Guideline
	The volume of new evidence	[4, 6]	Guideline

The nature of new evidence	[4]	Guideline
The quality of new evidence	[6]	Guideline
Likelihood of additional studies influencing current findings	[9]	Guideline
New evidence that would significantly modify recommendations	[19]	Guideline
How much new evidence is available	[32]	Guideline
Does the strength of the original clinical practice guideline recommendations remain the same?	[55]	Guideline
Does this new information significantly affect recommendations?	[55]	Guideline
Is the newly published information likely to change the grade of recommendation in the existing guideline?	[8]	Guideline
Is the newly published information likely to reverse a recommendation?	[8]	Guideline
Do direct consequences arise from the new potentially relevant evidence (e.g. change of recommendation or new recommendation)?	[15]	Guideline
Large, randomized placebo-controlled trial(s)	[9]	Guideline
New scientific insights based on literature research [Translated from Dutch: Nieuwe wetenschappelijke inzichten, vast te stellen op basis van literatuuronderzoek]	[29]	Guideline
Number of participants in new studies	[21]	Systematic review
Change in width of 95% confidence interval	[49]	Systematic review
Change in statistical significance	[7, 16, 33, 38, 39, 46, 48, 49]	Systematic review
Change in effect size	[7, 16, 33, 38, 39, 46, 48, 49]	Systematic review
Change in clinical significance given specified limits of clinical equivalence	[52]	Systematic review
Will new studies/information/data change findings or credibility?	[20]	Systematic review
Number of new studies identified	[21]	Systematic review
Heterogeneity in the out of date meta-analysis based on tau squared	[52]	Systematic review
Number of new relevant trials	[3]	Systematic review
Months since completion of the original review or last full update	[3]	Systematic review

	Totality (comprehensiveness) of all new evidence or data including harms & benefits	[21]	Systematic review
Access and inequality	Emergence of any evidence of inequality in access to services between different social groups that can be addressed through guideline recommendations	[41]	Guideline
	Evidence of impacts on equality groups	[41]	Guideline
	Extent to which the guideline is used in daily practice	[32]	Guideline
	Has a review had a good access or use?	[20]	
Resources	Changes in the restitution of diagnosis and treatment [Translated from Dutch: Veranderingen in de vergoeding van diagnostiek en behandeling]	[29]	Guideline
	Changes in the available resources [Translated from Dutch: Verandering in de beschikbare middelen]	[29]	Guideline
Organizational decisions	Impact on current and/or likelihood of need to develop new performance measure(s)	[9]	Guideline
	Seriousness of the problems or barriers to the implementation of the guideline	[32]	Guideline
	Need for consistency with a new guideline or guideline revision	[9]	Guideline
	Need for an internal organizational decision	[21]	Systematic review
Credibility	Time credibility	[21]	Systematic review
	Will adoption of new methods change findings or credibility?	[20]	Systematic review

Table A7 – Examples from literature review which could be used as staging tests

Domain	Item	Reference	Described for
Traits of evidence	Number of previous trials showing consistent results	[9]	Guideline
	The ratio of the observed number of additional participants to the predicted number of additional participants to obtain statistical significance	[12]	Systematic review
	The probability of producing statistically significant results when adding further studies (in which the results are consistent with those that already exist) to an existing meta-analysis	[51]	Systematic review
	The number of additional subjects (on average) to obtain a statistically significant result from a null meta-analysis. (Barrowman's n, the "new participant:n ratio")	[51]	Systematic review

	The ratio of the standard error of the predicted new estimate of effect to the existing effect	[51]	Systematic review
	The ratio of the sum of weights allotted to the predicted new and existing studies in the updated meta-analysis	[51]	Systematic review
	Study ratio where the total number of studies in the updated meta-analysis is compared to the total number of new studies.	[52]	Systematic review
	Estimated probability of conclusions changing after the addition of new studies to an existing meta-analysis, by using: estimated p = invlogit(0.1207 + 0.4101 x weight ratio + 0.1836 x number of new trials)	[52]	Systematic review
	Increase in number of patients	[49]	Systematic review
	The likelihood of the effect size of the updated meta-analysis to lay inside, outside, or across prespecified limits representing clinical equivalence	[51]	Systematic review
Expert and user solicitation	Opinions expressed by experts and by clinical practice guideline authors: are there results which were once considered important but no longer?	[55]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: should any diagnostic or treatment procedures be suspended or replaced by other procedures?	[55]	Guideline
	User perceptions: have any treatment preferences/consequences been identified in connection with the clinical practice guideline recommendation?	[55]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: should new recommendations within the scope of the original clinical practice guideline be included?	[55]	Guideline
	Other sources of information on the continued relevance of the guideline	[36]	Guideline
	The views of the Committee and topic experts.	[36]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: does the new evidence alter the risk/benefit ratio?	[55]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: is the new evidence significant enough to invalidate CPG recommendations?	[55]	Guideline
	Identification and assessment of new evidence: have literature limited searches identified new evidence that invalidates the CPG recommendations?	[55]	Guideline
	Opinions expressed by experts and by clinical practice guideline authors: are there any data showing that clinical practice is appropriate and the CPG is no longer necessary?	[55]	Guideline
Other	Data from visitations, performance measures, or other registries [Translated from Dutch: Gegevens uit visitatie, indicatoren of andere registraties]	[29]	Guideline
	Other information relevant to the updating of guidelines may be obtained from evaluation studies conducted to examine the implementation of effects of the guideline; from development and use of guideline-derived medical review criteria, performance measures, and standards of quality; or from other related activities.	[6]	Guideline

Table A8 – Examples from the literature review which could be used as staging thresholds

Domain	Item	Reference	Described for
Traits of evidence	Trial with sample size at least 3 times the size of previous largest trial	[49]	Systematic review
	A new study with at least three times the number of participants as in previous studies	[16, 33, 46]	Systematic review
	Change of at least 50% (for variable: change in width of 95% confidence interval)	[49]	Systematic review
	Change of least 50% (for variable: change in effect size)	[7, 16, 33, 38, 39, 46, 48, 49]	Systematic review
	New and old point estimates differ significantly	[48]	Systematic review
	At least 50% (for test: increase in number of patients)	[49]	Systematic review
	Change in statistical significance at a specified alpha level (alpha = 0.05)	[52]	Systematic review
	A probability of around 50% is suggestive of the need to update the SR, but any threshold can be chosen (for test: Estimated probability of conclusions changing after the addition of new studies to an existing meta-analysis, by using: estimated p = invlogit(0.1207 + 0.4101 x weight ratio + 0.1836 x number of new trials))	[52]	Systematic review

Table A9 – Examples from the literature review which could be used as management indications

Domain	Items	Reference	Described for
Changes in the evidence	Major changes in the literature may necessitate more frequent updates	[10]	Guideline
	Whether the new evidence provides data about specific uncertainties expressed in the original guidance	[4]	Guideline
	Does new scientific literature indicate an update is needed?	[27]	Guideline
	Potential key references (for a clinical question) that could potentially trigger an update	[30]	Guideline
	Recommendations with one or more key references that could potentially trigger an update.	[31]	Guideline

Potentially invalidating changes in evidence due to: Opposing findings	[7, 16, 33, 38, 46, 48, 49]	Systematic review
Potentially invalidating changes in evidence due to: Substantial harm	[7, 16, 33, 38, 46, 48, 49]	Systematic review
Potentially invalidating changes in evidence due to: Superior new treatment	[7, 16, 33, 38, 46, 48, 49]	Systematic review
Major changes in evidence due to: Important changes in effectiveness short of 'opposing findings'	[7, 16, 33, 38, 46, 48, 49]	Systematic review
Major changes in evidence due to: Expansion of treatment	[48, 49]	Systematic review
Major changes in evidence due to: Clinically important expansion of treatment	[7, 16, 33, 38, 46]	Systematic review
Major changes in evidence due to: Important caveat	[48, 49]	Systematic review
Major changes in evidence due to: Clinically important expansion of treatment	[7, 16, 33, 38, 46]	Systematic review
Major changes in evidence due to: Opposing findings from discordant meta-analysis or non-pivotal trial	[7, 16, 38, 46, 48, 49]	Systematic review
Major changes in evidence due to: Opposing findings from discordant pivotal trial or systematic review/ meta-analysis	[33]	Systematic review
New evidence in a situation where the original report had no evidence	[16, 33]	Systematic review
A major increase in the number of new studies	[16, 33, 46]	Systematic review
Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook does not contradict the previous review, but characterizes benefit in substantially different terms	[7, 38, 46]	Systematic review
Pivotal trial, meta-analysis including at least one new trial, practice guideline, recent textbook calls into question the use of the treatment on the basis of harm. A new result for harm that does not undermine use altogether but has clear potential to affect clinical decision making would count as a 'major change'.	[48, 49]	Systematic review
A pivotal trial or systematic review (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision-making.	[7, 16, 33, 38, 46, 48, 49]	Systematic review

	A pivotal trial or meta-analysis (or guideline) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.	[16, 33]	Systematic review
	Pivotal trial, systematic review including at least one new trial, practice guideline, or recent textbook characterized another treatment as significantly superior to the one evaluated in the original meta-analysis (based on efficacy or harm)—to the point that it would be preferred in most settings.	[48, 49]	Systematic review
	A pivotal trial or systematic review (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.	[7, 38, 46]	Systematic review
	A pivotal trial or meta-analysis (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.	[16, 33]	Systematic review
	Pivotal trial, new metaanalysis, more recent practice guideline, or recent textbook does not contradict the previous review, but characterizes benefit in substantially different terms	[46, 48, 49]	Systematic review
	Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook has expanded of the role of the treatment	[48, 49]	Systematic review
	Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook adds an important caveat, about the patient populations who benefit, way in which treatment has to be delivered in order to derive benefit, sustainability of benefit, or increases in harm that are not sufficient to undermine use altogether, but would clearly affect the decision to recommend treatment for at least some patient populations.	[48, 49]	Systematic review
	The treatment has been characterized in sufficiently different terms to the cohort review that disagreement would have met criteria for 'opposing findings' except the source was not a pivotal trial, new meta-analysis, or more recent practice guideline, or recent textbook	[48, 49]	Systematic review
Impact on questions, findings, conclusions, or recommendations	A specific issue such as a new drug therapy or national issue such as a new government policy will give rise to a new key question	[41]	Guideline
	The standard of care has shifted significantly since the last version of the document such that the questions only address the topic in part	[5]	Guideline
	There are new significant options [for treatment, diagnosis, etc.] available that are not covered by the current questions, such that new questions would need to be added to the document	[5]	Guideline
	New evidence relating to the topic that may warrant inclusion of additional recommendations.	[25]	Guideline
	New evidence substantially changes a small number of recommendations in the guideline (corresponding to no more than two related key questions)	[41]	Guideline
	The new evidence is inconsistent with the data used to inform the original practice guideline report. The strength of the new evidence will alter the conclusions of the original document. Recommendations in the original report will change.	[28]	Guideline

	The new evidence is consistent with the data used to inform the original practice guideline report. The strength of the recommendation in the original report has been modified to reflect this additional evidence.	[28]	Guideline
	Emerging data that suggests that the current recommendations need to be revised. (For rapid advice guidelines in case of a public health emergency)	[56]	Guideline
	On initial review, does the newly identified evidence support the existing recommendation?	[5]	Guideline
	New evidence relating to the topic that may warrant removal or change to a recommendation or suggestion	[25]	Guideline
	New evidence relating to the topic that may alter the strength of a recommendation	[25]	Guideline
	The availability of new methods that would have a meaningful impact on the review findings.	[17]	Systematic review
	The availability of new data that would have a meaningful impact on the review findings.	[17]	Systematic review
	Some new evidence that might change the comparative effectiveness review's conclusion, and /or a minority of responding experts assessed the review's conclusion as having new evidence that might change the conclusion	[7, 38, 42, 44-46]	Systematic review
	Substantial new evidence that might change the comparative effectiveness review's conclusion, and/or a majority of responding experts assessed the review's conclusion as having new evidence that might change the conclusion	[7, 38, 42, 44-46]	Systematic review
	New evidence that rendered the comparative effectiveness review's conclusion out of date or no longer applicable	[7, 38, 42, 44-46]	Systematic review
Organizational decisions	The revised guideline can be included in the short term in a multidisciplinary guideline/care standard	[32]	Guideline
	The revised guideline may be important for positioning the health care discipline	[32]	Guideline
	Did this guideline topic spur additional high-quality research from the specialty societies or the organization's membership?	[8]	Guideline
	Topic is otherwise outside the organization's scope	[53]	Guideline
	Inclusion in policy decision making or clinical practice guidelines (e.g. it might be important to update a review to include it in a new clinical guideline. If any such factors (termed updating signals) are identified, then a judgement is made on whether a signal for updating is likely or unlikely to change the results or conclusions of the review)	[52]	Systematic review
	Inclusion in policy or clinical guidelines	[54]	Systematic review
Implementation	Evidence that the guideline is fully complied with by relevant organizations, and has become accepted practice	[41]	Guideline

	There is evidence that the CPG has been fully implemented in the healthcare system and has been accepted as clinical practice	[55]	Guideline
	Existing controversy or gap between evidence and practice	[53]	Guideline
	Topic is not relevant to primary care provider because the service is not implemented in a primary care setting or not referable by a primary care provider	[53]	Guideline
Superseded or outdated	Not revised or amended following 10 years of publication	[11]	Guideline
	The recommendations no longer apply, but the guideline is not of sufficiently high priority for updating.	[36]	Guideline
	Topic is no longer relevant to clinical practice because of changes in technology, new understanding of disease etiology/natural history, or evolving natural history of the disease	[53]	Guideline
	Discovery of a new preventive or treatment measures make the CPG obsolete	[55]	Guideline
	The CPG recommendations are no longer applicable or are outdated	[55]	Guideline
	Emergence of new treatments or preventive measures that render the guideline irrelevant	[41]	Guideline
	Superseded by a more recent or more comprehensive guideline	[41]	Guideline
	If a therapy was no longer favored, no longer in use, or in question because of a safety concern	[38]	
Postpone	The new evidence is inconsistent with the data used to inform the original practice guideline report. However, the strength of the new evidence does not alter the conclusions of the original document.	[28]	Guideline
	The new evidence is consistent with the data used to inform the original practice guideline report. The recommendations in the original report remain unchanged.	[28]	Guideline
	Is there a good reason (e.g. new stronger evidence will be published soon, changes to current recommendations are trivial or address very limited situation) to postpone updating the guideline?	[5]	Guideline
	Should full assessment and review of this document be deferred until next year? Consider yes if the document is less than three years old and there is no reason to doubt the recommendations	[5]	Guideline
	Should full assessment and review of this document be deferred until next year? Consider yes if the document is between three and five years old and a justification can be provided as to why the recommendations can be considered trustworthy for another year	[5]	Guideline
Other	In general, if you believe that for the document to still be useful it will have to substantially be rewritten	[5]	Guideline
	The document has been repeatedly deferred, and is now older than 5 years	[5]	Guideline
	Guidelines that contain 50% or greater of its recommendations as consensus (expert opinion) or inconclusive generally should not be updated.	[8]	Guideline
	Topic had a low public health burden	[53]	Guideline

	The subject of another more recent CPG partly or wholly overlaps with that of the CPG	[55]	Guideline
	Continued uncertainty regarding the study intervention	[50]	Systematic review

Table A10 – Empty process description table

Diagnosis	*Target condition definition*	<definition>	
	Detection variable	<variable>	
	Detection test (protocol)	<test> <protocol>	
	Detection test threshold	<threshold>	
		Target condition present	**Target condition absent**
Staging	*Staging variable*	<variable>	<variable>
	Staging test (protocol)	<test> <protocol>	<test> <protocol>
	Staging thresholds	<thresholds>	<thresholds>
		Target condition present	**Target condition absent**
Management	*Management options (indications) and actions*	<option 1> <indications> <actions>	<option 1> <indications> <actions>
		<option …> <indications> <actions>	<option …> <indications> <actions>
Monitoring	*Monitoring*	<timeframe>	

Table A11 – Process description table of the example strategy in organization A

Diagnosis	*Target condition definition*	A CPG recommendation is outdated when new evidence is present		
	Detection variable	New peer-reviewed scientific evidence published since the publication of the CPG		
	Detection test (protocol)	A MEDLINE, EMBASE, and Cochrane Library search (A repetition of the original MEDLINE, EMBASE, and Cochrane Library search is conducted. A double-blind title/abstract and full-text selection is performed based on the original selection criteria)		
	Detection test threshold	Any newly identified relevant peer-reviewed scientific evidence will result in the presence of the target condition		
		Target condition present		**Target condition absent**
Staging	*Staging variable*	Likelihood of potential changes in the strength of the body of evidence *(target condition present)*	Number of new studies *(target condition present)*	Relevance for clinical practice (target condition absent)
	Staging test (protocol)	Informal GRADEing of the body of evidence (The new studies are added to the existing body of evidence. GRADE is informally used to assess the likelihood that the addition of new evidence results in a change of strength.)	A MEDLINE, EMBASE, and Cochrane Library search (A repetition of the original MEDLINE and Cochrane Library search is conducted. A double-blind title/abstract and full-text selection is performed based on the original selection criteria. The number of selected studies is counted.)	A guideline panel vote (The guideline panel is asked whether the recommendation is still relevant for clinical practice. The panel answers dichotomously [yes/no]. The proportion indicating that there is no longer any relevance is used for staging.)
	Staging thresholds	Likely change of strength: Any probable change in the strength (according to the GRADE certainty of evidence) No likely change of strength: No probable change in the strength (according to the GRADE certainty of evidence).	Low impact: 1 study Moderate impact: 2-3 studies High impact: ≥4 studies	Still relevant: 0-30% of the panel indicate that there is no relevance Unsure: 30-60% of the panel indicate there is no relevance No longer relevant: >60% of the panel indicate there is no relevance

		Target condition present	Target condition absent
Management	*Management options (indications) and actions*	Update (likely change of strength OR high impact of the number of new studies) Actions: Initiate an update, inform guideline panel, allocate resources	Do not update, reassess at a later time point (still relevant or unsure) Actions: Return the recommendation to the portfolio. Reassess at a later point in time.
		Do not update, reassess at a later point in time (no likely change of strength) Actions: Return the recommendation to the portfolio. Reassess at a later point in time.	Withdraw (no longer relevant) Actions: Withdraw the recommendation from the guideline (and portfolio). Inform the guideline panel. Inform the end-users with a notice on the website.
Monitoring	*Monitoring*	The recommendations are assessed every year	

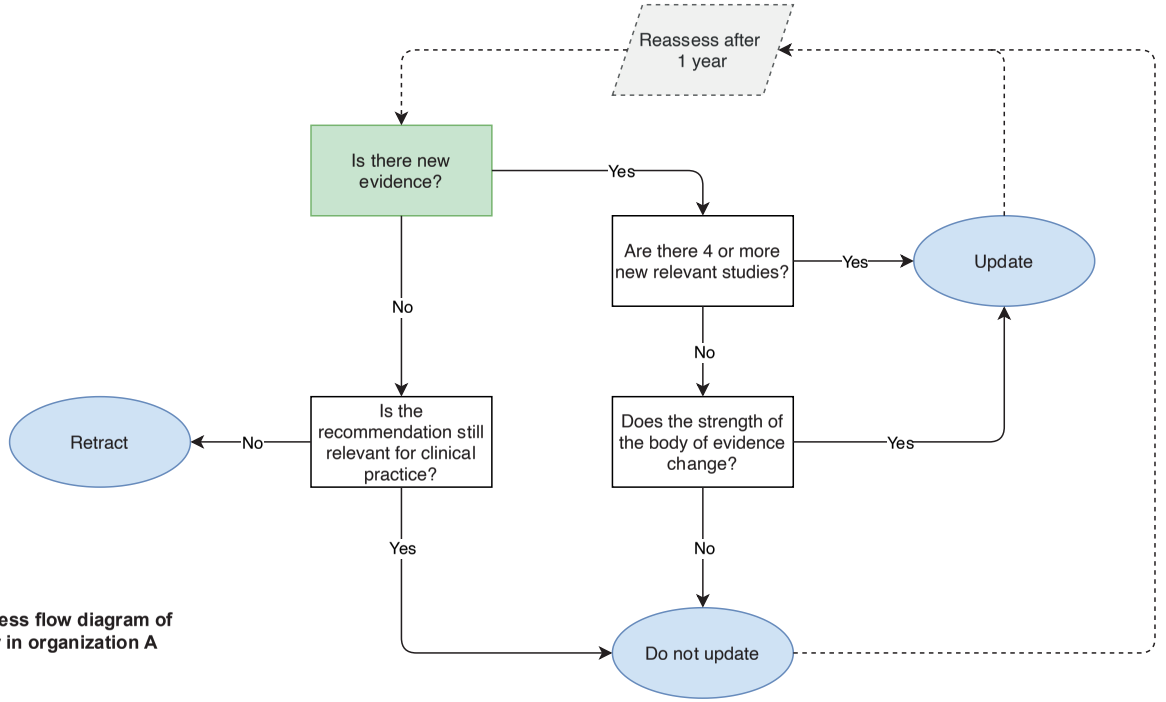


Figure A3 (left). Process flow diagram of the example strategy in organization A *(see Table A11)*

Table A12 – Process description table of the example strategy in organization B

Diagnosis	*Target condition definition*	A CPG recommendation is outdated when new interventions are available or when new evidence is present.		
	Detection variable	Availability of new intervention	Newly identified evidence	
	Detection test (protocol)	Survey among experts to identify newly available interventions (Experts are identified. A survey is prepared and sent to the experts. Experts can indicate whether they believe there are new intervention available.)	Survey among experts to identify newly available evidence (Experts are identified. A survey is prepared and sent to the experts. Experts can suggest new evidence.)	A limited literature search in the top 5 general topic journals and top 5 specialty journals (The top 5 general and specialty journals were ranked on impact factor. A specific search was conducted in MEDLINE, limited to the selected journals. Literature selection was performed by a methodologist.)
	Detection test threshold	The identification of at least one new intervention will result in the presence of the target condition. Relevance is defined as being a preventive intervention resulting in the abundance of the current interventions or as an alternative to the current interventions.	Any new relevant evidence identified by the experts will result in the presence of the target condition. Relevance is defined by the original selection criteria.	Any new evidence identified from the literature search will result in the presence of the target condition.

		Target condition present			**Target condition absent**
Staging	*Staging variable*	Number of new relevant interventions	Relevance of newly identified interventions	Quality of new studies	Unwanted practice variation
	Staging test (protocol)	Survey among experts to count newly available interventions (Experts are identified. A survey is prepared and sent to the experts. Experts can indicate whether they believe there are new intervention available. Relevance is scored by the guideline panel.)	A survey for the guideline panel to assess relevance (The guideline panel is consulted to assess the relevance of the identified interventions by the expert survey. Relevance is scored on a 7-point likert-scale: 0 = no relevante / 6 = highly relevant.)	Risk of bias assessment (A methodologist assesses the risk of bias for newly identified potentially relevant RCTs with the Cochrane Risk of Bias Tool 2.0)	National data-registry analysis (The recommendation is assessed according to the data from a national data-registry. Practice variation is assessed within the population.)

	Staging thresholds	New interventions available: Any new intervention was identified No new interventions available: No new intervention was identified	No relevance: A mean score of >1 Low relevance: A mean score of 1-3 Some relevance: A mean score of 3-5 High relevance: A mean score of >5 points	High quality studies available: At least one RCT with low risk of bias is available Only low quality studies available: Only RTCs with high or unclear risk of bias are available	No variation: The recommendation is usually appropriately applied in practice indicating no further guidance is needed No unwanted variation: There are signs that there are deviations from the recommendation in clinical practice that can be considered normal Unwanted variation: There are signs that there are deviations from the recommendation in clinical practice that cause harm

		Target condition present	**Target condition absent**
Management	Management options (indications) and actions	Update ((new intervention available AND some or high relevance of the intervention) OR high quality study available OR unwanted practice variation) Actions: Initiate an update, inform guideline panel, allocate resources	Do not update, reassess at a later point in time (no unwanted variation) Actions: Return the recommendation to the portfolio. Reassess at a later point in time.
		Do not update, reassess at a later point in time (No new relevant intervention available AND only low quality study available AND no variation or no unwanted variation) Actions: Return the recommendation to the portfolio. Reassess at a later point in time..	Archive (no variation) Actions: Keep the recommendation on the website. Include a note that the recommendation is archived and is not to be updated.
			Update (unwanted variation) Actions: Initiate an update, inform guideline panel, allocate resources
Monitoring	*Monitoring*	The recommendations are assessed every two years.	

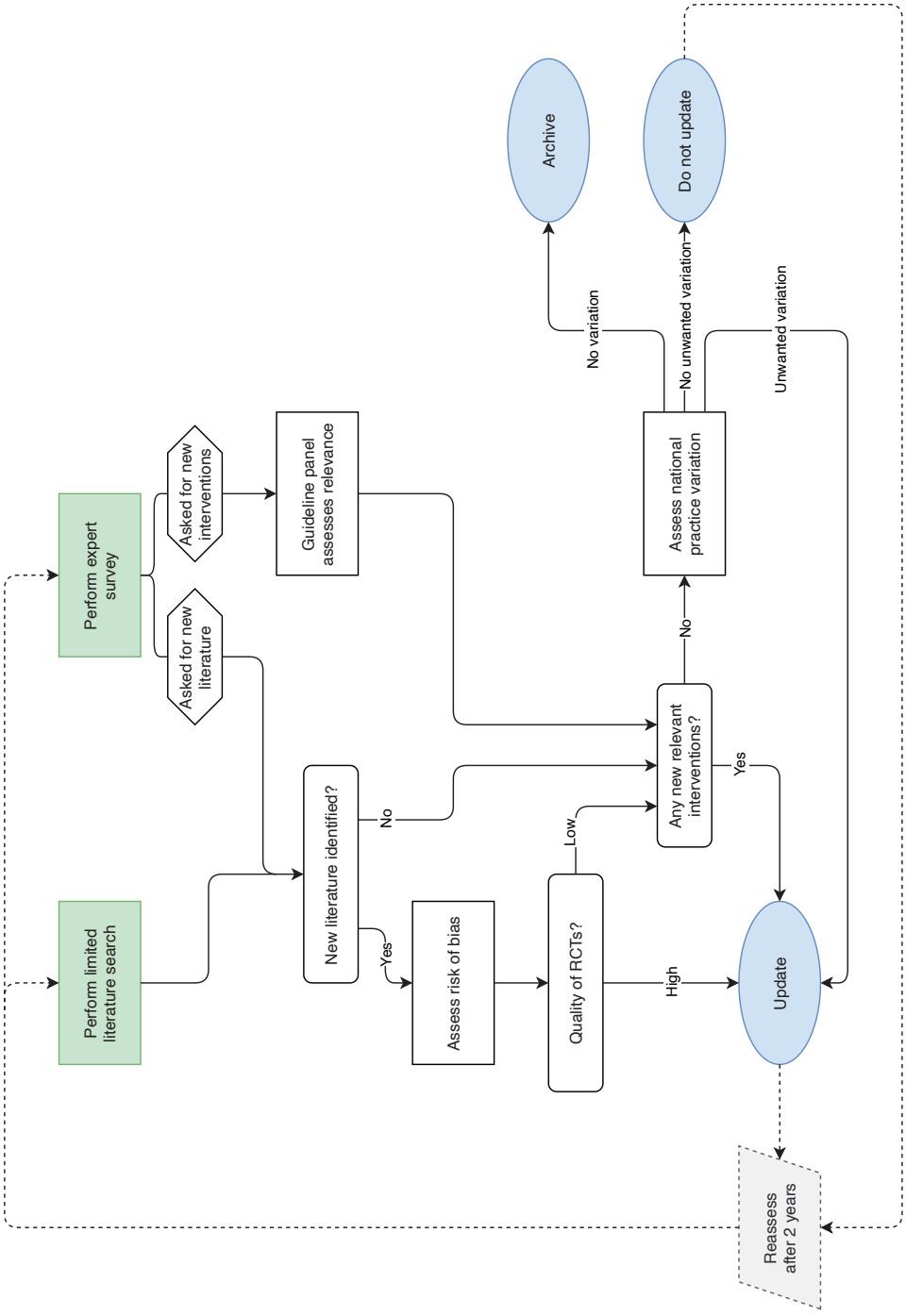


Figure A4. Process flow diagram of the example strategy in organization B *(see Table A12)*

Table A13 – A hypothetical example of a 'living' recommendations strategy based on information found in Akl et al. 2017 [57], El Mikati et al. 2022 [58], and Bragge et al. 2022 [59] which we adapted for illustrative purposes

Diagnosis	*Target condition definition*	The guideline recommendation is out of date when any new evidence becomes available		
	Detection variable	New scientific evidence according to the existing literature selection criteria		
	Detection test (protocol)	Literature search and selection (MEDLINE and EMBASE are searched by rerunning the previous search strategy. The search strategy was developed by an information specialist. The title and abstract of the retrieved hits are screened by two methodologists independently. Conflicts are resolved. The resulting full text articles are screened by two methodologists for inclusion. Data is extracted from the Included articles and study quality is appraised.)		
	Detection test threshold	Any new relevant published evidence		
		Target condition present		**Target condition absent**
Staging	*Staging variable*	Impact of new relevant published evidence on the recommendations	Impact of new relevant published evidence on the certainty of evidence	–
	Staging test (protocol)	Expert panel discussion and voting (The panel is supplied with the review question and guideline recommendation, with the new evidence and study quality, and with a set of questions to determine the influence of the new evidence on the recommendation. The panel discusses the possible implications in a physical meeting and votes whether the recommendation is impacted by the new evidence)	Certainty of evidence evaluation (The GRADE Summary of Findings table is updated by methodologists with information from the new evidence and performing new GRADE assessments)	–
	Staging thresholds	**Impacted:** when >75% of the panel indicates that the new evidence has impact on the recommendation **Uncertain impact:** 25-75% of the panel indicates that the new evidence has impact on the recommendation **No impact:** >25% of the panel indicates that the new evidence has impact on the recommendation	**High certainty:** when all critical outcomes for decision-making have a HIGH GRADE. **Relatively certain:** when most, but not all, critical outcomes for decision-making have HIGH GRADE **Moderate certainty:** when most or all critical outcomes for decision-making have MODERATE GRADE.	–

		Uncertain: When most critical outcomes for decision-making have LOW or VERY LOW GRADE	

		Target condition present	Target condition absent
Management	*Management options (indications) and actions*	**Update the recommendation and continue monitoring** (when the expert panel indicates the recommendation is impacted by new evidence [>75%]) Perform an update using the regular procedures	**Do not update and continue monitoring** (When no new relevant evidence was found) Continue monitoring as defined under 'Monitoring'
		Do not update and continue monitoring (When the expert panel was uncertain [25-75%] OR indicated that there was no impact [<25%]) Continue monitoring as defined under 'Monitoring'	**–**
		Update the recommendation and retire from living status (when the expert panel indicates the recommendation is impacted by new evidence [>75%] AND the certainty of evidence became high) Perform an update using the regular procedures and retire the recommendation from participating in this maintenance strategy.	–
Monitoring	*Monitoring*	Weekly reruns of the existing search strategy	

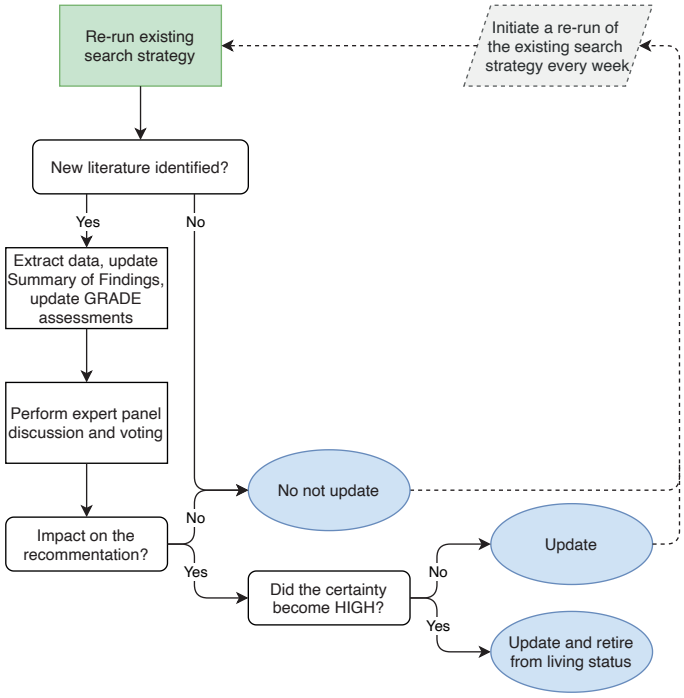


Figure A5 – Process flow diagram of the hypothetical 'living' example strategy *(see Table A13) based on information found in Akl et al. 2017 [57], El Mikati et al. 2022 [58], and Bragge et al. 2022 [59] which we adapted for illustrative purposes.*

LITERATURE LIST

1. PubMed Health to be discontinued October 31, 2018; content will continue to be available at NLM [https://ncbiinsights.ncbi.nlm.nih.gov/2018/07/31/pubmed-health-discontinued-october-2018-nlm/]
2. Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L: Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. Implement Sci 2014, 9:3.
3. Peterson K, McDonagh MS, Fu R: Decisions to update comparative drug effectiveness reviews vary based on type of new evidence. J Clin Epidemiol 2011, 64(9):977-984.
4. Lyratzopoulos G, Barnes S, Stegenga H, Peden S, Campbell B: Updating clinical practice recommendations: is it worthwhile and when? Int J Technol Assess Health Care 2012, 28(1):29-35.
5. Agbassi C, Messersmith H, McNair S, Brouwers M: Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. J Clin Epidemiol 2014, 67(12):1335-1342.
6. Agency for Health Care Policy and Research: Process for determining need for updates of clinical practice guidelines. Fed Regist 1994, 59(79 Pt 1):19723-19725.
7. Ahmadzai N, Newberry SJ, Maglione MA, Tsertsvadze A, Ansari MT, Hempel S, Motala A, Tsouros S, Schneider Chafen JJ, Shanman R et al: A surveillance system to assess the need for updating systematic reviews. Syst Rev 2013, 2:104.
8. American Academy of Orthopaedic Surgeons Clinical Practice Guideline Unit: Guideline Process Procedure for Updating Existing Evidence-Based Clinical Practice Guidelines. In., edn.: American Academy of Orthopaedic Surgeons; 2011.
9. American College of Cardiology Foundation and American Heart Association: Methodology manual and policies from the ACCF/AHA Task Force on Practice Guidelines. In., edn.; 2010.
10. American College of Occupational and Environmental Medicine: Methodology for ACOEM's Occupational Medicine Practice Guidelines - 2017 Revision. In., edn.; 2017.
11. American Urological Association: American Urological Association Clinical Practice Guidelines Development. In., edn.; 2015.
12. Barrowman NJ, Fang M, Sampson M, Moher D: Identifying null meta-analyses that are ripe for updating. BMC Med Res Methodol 2003, 3:13.
13. Bashir R, Surian D, Dunn AG: Time-to-update of systematic reviews relative to the availability of new evidence. Syst Rev 2018, 7(1):195.
14. Becker M, Neugebauer EA, Eikermann M: Partial updating of clinical practice guidelines often makes more sense than full updating: a systematic review on methods and the development of an updating procedure. J Clin Epidemiol 2014, 67(1):33-45.
15. Becker M, Jaschinski T, Eikermann M, Mathes T, Bühn S, Koppert W, Leffler A, Neugebauer E, Pieper D: A systematic decision-making process on the need for updating clinical practice guidelines proved to be feasible in a pilot study. J Clin Epidemiol 2018, 96:101-109.
16. Chung M, Newberry SJ, Ansari MT, Yu WW, Wu H, Lee J, Suttorp M, Gaylor JM,

Motala A, Moher D et al: Two methods provide similar signals for the need to update systematic reviews. J Clin Epidemiol 2012, 65(6):660-668.
17. Cumpston M, Chandler J: Chapter IV: Updating a review. In: Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021). edn. Edited by Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA: Cochrane; 2021.
18. Davis DA, Goldman J, Palda VA, Canadian Medical Association: Handbook on clinical practice guidelines. Ottawa: Canadian Medical Association = Association médicale canadienne; 2007.
19. Dumonceau JM, Hassan C, Riphaus A, Ponchon T: European Society of Gastrointestinal Endoscopy (ESGE) Guideline Development Policy. Endoscopy 2012, 44(6):626-629.
20. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, Beyene J, Chang S, Churchill R, Dearness K et al: When and how to update systematic reviews: consensus and checklist. Bmj 2016, 354:i3507.
21. Garritty C, Tsertsvadze A, Tricco AC, Sampson M, Moher D: Updating systematic reviews: an international survey. PLoS One 2010, 5(4):e9914.
22. Guidelines and Protocols Advisory Committee: Guidelines and Protocols Advisory Committee Handbook: How our "Made in BC" Clinical Practice Guidelines and Protocols are Developed. In., edn.; 2017.
23. Haller MC, van der Veer SN, Nagler EV, Tomson C, Lewington A, Hemmelgarn BR, Gallagher M, Rocco M, Obrador G, Vanholder R et al: A survey on the methodological processes and policies of renal guideline groups as a first step to harmonize renal guidelines. Nephrol Dial Transplant 2015, 30(7):1066-1074.
24. Haute Autorité de Santé: Élaboration de recommandations de bonne pratique: Méthode Recommandations pour la pratique clinique. In., edn. Saint-Denis La Plaine: HAS; 2016.
25. Howell M: KHA-CARI guidelines development manual. In., edn.; 2015.
26. Iorio A, Ageno W, Cosmi B, Imberti D, Lussana F, Siragusa S, Tormene D, Tosetto A, Cattaneo M: Objectives and methodology: Guidelines of the Italian Society for Haemostasis and Thrombosis (SISET). Thromb Res 2009, 124(5):e1-5.
27. Javaher SP: Guideline Development Process in a Public Workers' Compensation System. Phys Med Rehabil Clin N Am 2015, 26(3):427-434.
28. Johnston ME, Brouwers MC, Browman GP: Keeping cancer guidelines current: results of a comprehensive prospective literature monitoring strategy for twenty clinical practice guidelines. Int J Technol Assess Health Care 2003, 19(4):646-655.
29. Kwaliteitsinstituut voor de Gezondheidszorg CBO: Evidence-based richtlijn ontwikkeling: Handleiding voor werkgroepen. In., edn.; 2007.
30. Martínez García L, Pardo-Hernández H, Sanabria AJ, Alonso-Coello P: Continuous surveillance of a pregnancy clinical guideline: an early experience. Syst Rev 2017, 6(1):143.
31. Martínez García L, Sanabria AJ, García Alvarez E, Trujillo-Martín MM, Etxeandia-Ikobaltzeta I, Kotzeva A, Rigau D, Louro-González A, Barajas-Nava L, Díaz Del Campo P et al: The validity of recommendations from clinical guidelines: a survival analysis. Cmaj 2014, 186(16):1211-1219.
32. Meerhof G, Heijblom K, Knoop J: The Royal Dutch Society for Physical Therapy

(KNGF) Guideline methodology manual 2016: Methodology for developing/updating and implementing KNGF guidelines. In., edn.; 2016.
33. Mickenautsch S, Yengopal V: The modified Ottawa method to establish the update need of a systematic review: glass-ionomer versus resin sealants for caries prevention. J Appl Oral Sci 2013, 21(5):482-489.
34. Murad MH, Montori VM, Sidawy AN, Ascher E, Meissner MH, Chaikof EL, Gloviczki P: Guideline methodology of the Society for Vascular Surgery including the experience with the GRADE framework. J Vasc Surg 2011, 53(5):1375-1380.
35. National Health and Medical Research Council: A guide to the development, implementation, and evaluation of clinical practice guidelines. 2009.
36. National Institute for Health and Care Excellence: Developing NICE guidelines: the manual (April 2017 update). In., edn.; 2014.
37. National Institute for Health and Care Excellence: NICE Process and Methods Guides. In: Interim Clinical Guideline Surveillance Process and Methods Guide 2013. edn. London: National Institute for Health and Care Excellence (NICE); 2013.
38. Newberry SJ, Ahmadzai N, Motala A, Tsertsvadze A, Maglione M, Ansari MT, Hempel S, Tsouros S, Schneider Chafen J, Shanman R et al: AHRQ Methods for Effective Health Care. In: Surveillance and Identification of Signals for Updating Systematic Reviews: Implementation and Early Experience. edn. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
39. Pattanittum P, Laopaiboon M, Moher D, Lumbiganon P, Ngamjarus C: A comparison of statistical methods for identifying out-of-date systematic reviews. PLoS One 2012, 7(11):e48894.
40. Rosenfeld RM, Shiffman RN, Robertson P: Clinical Practice Guideline Development Manual, Third Edition: a quality-driven approach for translating evidence into action. Otolaryngol Head Neck Surg 2013, 148(1 Suppl):S1-55.
41. Scottish Intercollegiate Guidelines Network: SIGN 50: a guideline developer's handbook. In: SIGN publication no 50. edn. Edinburgh: SIGN; 2015.
42. Shekelle PG, Motala A, Johnsen B, Newberry SJ: Assessment of a method to detect signals for updating systematic reviews. Syst Rev 2014, 3:13.
43. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, Woolf SH: Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? Jama 2001, 286(12):1461-1467.
44. Shekelle P, Newberry S, Maglione M, Shanman R, Johnsen B, Carter J, Motala A, Hulley B, Wang Z, Bravata D et al: AHRQ Methods for Effective Health Care. In: Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005–2009). edn. Rockville (MD): Agency for Healthcare Research and Quality (US); 2009.
45. Shekelle PG, Motala A, Johnsen B: AHRQ Methods for Effective Health Care. In: Assessment of a Method to Detect Signals for Updating Systematic Reviews. edn. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
46. Shekelle PG, Newberry SJ, Wu H, Suttorp M, Motala A, Lim YW, Balk EM, Chung M, Yu WW, Lee J et al: AHRQ Methods for Effective Health Care. In: Identifying Signals for Updating Systematic Reviews: A Comparison of Two Methods. edn. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.
47. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH: When should clinical guidelines be

updated? Bmj 2001, 323(7305):155-157.
48. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D: How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 2007, 147(4):224-233.
49. Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, Moher D: AHRQ Technical Reviews. In: Updating Systematic Reviews. edn. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007.
50. Soll RF: Updating reviews: the experience of the Cochrane Neonatal Review Group. Paediatr Perinat Epidemiol 2008, 22 Suppl 1:29-32.
51. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A: An encouraging assessment of methods to inform priorities for updating systematic reviews. J Clin Epidemiol 2009, 62(3):241-251.
52. Takwoingi Y, Hopewell S, Tovey D, Sutton AJ: A multicomponent decision tool for prioritising the updating of systematic reviews. Bmj 2013, 347:f7191.
53. U.S. Preventive Services Task Force: US Preventive Services Task Force Procedure Manual. In., edn.; 2015.
54. Welsh E, Stovold E, Karner C, Cates C: Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach. J Clin Epidemiol 2015, 68(3):341-346.
55. Working Group for CPG Updates: Updating Clinical Practice Guidelines in the National Health System: Methodology Handbook. In: Clinical Practice Guidelines in the National Health System: I+CS No2007/02-01. edn.: National Health System Quality Plan of the Spanish Ministry of Health and Social Policy, Aragon Health Sciences Institute (I+CS); 2009.
56. World Health Organization: WHO handbook for guideline development - 2nd edition. In., edn. Geneva, Switzerland: WHO Press; 2014.
57. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ: Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol 2017, 91:47-53.
58. El Mikati IK, Khabsa J, Harb T, Khamis M, Agarwal A, Pardo-Hernandez H, Farran S, Khamis AM, El Zein O, El-Khoury R et al: A Framework for the Development of Living Practice Guidelines in Health Care. Ann Intern Med 2022, 175(8):1154-1160.
59. Bragge P, Allison DJ, Mehta S, Guy S, Loh E: A pilot living clinical practice guidelines approach was feasible and acceptable to guideline panel members. J Clin Epidemiol 2022, 143:22-29.

CHAPTER 5

Priority-setting for resource constraints in systematic review or clinical practice guideline maintenance strategies: Extending the Portfolio Maintenance by Test-Treatment (POMBYTT) framework

M.S. Oerbekke
R.G. Elbers
M.J. van der Laan
L. Hooft

Submitted

ABSTRACT

BACKGROUND

In maintenance strategies for clinical practice guidelines and systematic reviews where resources are limited, priority-setting plays a crucial role. We aimed to extend the Portfolio Maintenance by Test-Treatment (POMBYTT) framework with guidance on priority-setting.

METHODS

A conceptual approach was employed combined with a literature review. Roles of new tests in a diagnostic work-up, as described in diagnostic rest accuracy methodology, were transferred to the context of systematic review and clinical practice guideline maintenance strategies.

RESULTS

The POMBYTT framework was extended with guidance incorporating priority-setting into maintenance strategies. Four distinct performances were identified based on the combination of the role (i.e. triage or add-on) and outcome level (i.e. ordinal or ranked continuous) of priority-setting assessments. The priority-setting assessment aims to assign a certain level of priority to systematic reviews or clinical practice guidelines. The priority-setting assessment determines whether the systematic review or clinical practice guideline continues within the maintenance strategy (triage) and whether the systematic review or clinical practice guideline receives a management action (add-on). The literature review provided possible examples of priority indicators for priority-setting assessments.

CONCLUSIONS

The performance of a priority-setting assessments should be aligned with the goal of priority-setting in the maintenance strategy. Selecting a specific performance will guide the designers of a maintenance strategy in determining the appropriate role and outcome level for the priority-setting assessment.

Keywords: priority-setting, prioritization, indicators, clinical practice guidelines as topic, systematic reviews as topic, concept formation

1. INTRODUCTION

Systematic review (SR) and clinical practice guideline (CPG) developing organizations usually encounter resource limitations in the production and updating of their products. Resources can be viewed in a broad sense where it could concern the available budget, personnel, or time. It is therefore important for the resource constrained organization to spend its resources on the maintenance of those SRs or CPGs which have enough priority. A priority-setting assessment can be incorporated in a maintenance strategy to identify which SR or CPG has enough priority to invest resources in. These assessments, for example the UpPriority tool,[1, 2] may incorporate priority indicators to assess the perceived priority of an SR or CPG, such as *"impact on access"* and *"known new evidence",*[3] or *"impact of outdated recommendations on safety"* and *"user's interest",*[1, 2] or *"uncertainty in practice"* and *"variation in practice".*[4, 5] Systematic reviews also identified various (steps in) priority-setting assessments and even more indicators.[6-9] Such assessments are used to identify priority SRs or CPGs, although their role and performance as part of a maintenance strategy is usually not made explicit. Priority-setting can be performed at different stages in a maintenance strategy. For example, during an updating strategy[3] or as a last step in the strategy before actually performing updates.[10] For living CPGs, it is even proposed to prioritize recommendations upfront to determine which should be assigned a living status.[11] Different timings may have different consequences for the maintenance strategy. Furthermore, identifying SRs or CPGs in need for updating might even be interpreted as being priority-setting itself.

The distinction between need for updating and priority-setting can be made clear through the Portfolio Maintenance by Test-Treatment (POMBYTT) framework.[12] The framework describes how SRs or CPGs undergo diagnostic and staging tests to determine which management option should be selected, for example whether an update is indicated. However, there may be more SRs or CPGs for which an update is indicated than there are resources to perform these updates. A priority-setting assessment is then needed to align the available resources to those SRs or CPGs with the highest priority to receive an update. Here, resources can be regarded in a broad sense. Previous work discussed considerations for the resource constrained organization when choosing tests and test variables in designing a maintenance strategy.[12] However, elements for priority-setting within a maintenance strategy for the resource constrained organization were not previously described in the POMBYTT framework. Such elements may contribute to explicitly thinking about the timing, performance, and indicators of priority-setting assessments in maintenance strategies. Extending the POMBYTT framework with priority-setting elements would help to design and understand how maintenance strategies operate under resource constraints. Other work may be used to design the priority-setting assessment itself, were various steps before, during and after priority-setting are described.[7]

The POMBYTT framework can be considered a theoretical approach describing a test-treatment pathway for maintenance. Adding priority-setting assessments can therefore be thought of as adding news tests to a test-treatment pathway. Interestingly, diagnostic test accuracy research methodology describes three roles for new tests to be added to an existing work-up, which are replacement, triage, and add-on.[13] These new tests will, however,

not function as diagnostic or staging tests. Where diagnostic and staging tests are used to determine which management option should be selected in the strategy, priority-setting assessments will test which SR or CPG has more priority above others. We therefore aimed to extend the POMBYTT framework by adding priority-setting elements concerning the role, performance, and examples of indicators to further aid resource constrained organizations in designing a maintenance strategy.

2. METHODS

We conducted a literature search to gain an overview of priority indicators described in the literature. Additional file 1 presents the detailed literature review methodology. In summary, we searched MEDLINE (through PubMed) and PubMed Health for relevant studies, reports, and handbooks. Two authors (RGE, MSO) independently conducted the title and abstract screening and the selection of the full text papers. One author (MSO) extracted the relevant data. We used NVivo for Windows (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12) for qualitative analysis to identify overall domains among the extracted data. We then transferred the roles of new tests described in diagnostic test accuracy methodology[13] to the context of priority-setting in the POMBYTT framework.[12] Three separate roles of diagnostic tests are defined: replacement of an existing test, triage before existing tests, and add-on after existing tests.[13] Replacement will result in using a new test on the exact same location of an existing test in the test-treatment pathway. Priority-setting assessments, however, will not replace any of the diagnostic or staging tests in the POMBYTT framework. Therefore, the replacement role is not further explored. Instead, we describe priority-setting in both the triage and add-on role. The extracted priority indicators from our literature review were mapped at our own discretion to the roles in the framework. In addition, we furthermore described how priority-setting assessments may perform in a maintenance strategy depending on their role and outcome level. Table 1 provides a glossary of terms used in the context of priority-setting in the POMBYTT framework.

3. RESULTS

3.1 PRIORITY INDICATORS AND OUTCOME LEVELS IN LITERATURE

Fifteen references were included in the literature review (see Figure A1 and Tables A1-2 in Additional file 1). Fifty priority indicators were described, such as *"Impact on access to health care",*[2] *"Are there substantial variations in the diagnosis and/or treatment of the health problem?",*[14] *"How out of date are the conclusion considering the magnitude/direction of changes in estimates",*[15] and *"Does the health problem in question carry a high individual or population burden of morbidity, mortality, or disability?".*[14] We identified nine domains of indicators through qualitative analysis: access, clinical practice, new evidence, prediction

Table 1 – Glossary of the terms used in the context of priority-setting in the POMBYTT framework.

Term	Description
Add-on	Add-on is one of the roles within priority-setting. In this role, the priority-setting assessment is placed after the initial staging and selecting the management option, but before any management actions. The add-on role ensures that priority systematic reviews and clinical practice guidelines receive the necessary management actions associated with the selected management option, while lower-priority systematic reviews and guidelines are deferred.
Cut-off	A cut-off refers to a predetermined limit that is used to select either specific levels (using an ordinal outcome level) or a specific number of the top-ranked systematic reviews or clinical practice guidelines (see limit of capacity).
Performance	The manner a priority-assessment operates within the maintenance strategy. The performance is affected by both the specific role the priority-setting assessment fulfills and its outcome level.
Limit of capacity	In priority-setting assessments with a ranked continuous outcome level, this specific type of cut-off used. This cut-off is aligned to the available resources that the organization is able or willing to spend. Let's consider an add-on priority-setting assessment that involves performing an update as an example. The limit of capacity is set at five, which means the organization has available resources guaranteeing five updates. As a result, the five top-ranked systematic reviews or clinical practice guidelines are selected to receive the associated management actions.
Maintenance strategy	This strategy aims to select the most appropriate management option through testing and performing specific management actions that are associated with the selected management option.
Management actions	Actions that follow from the decision for a specific management option. It may concern actions to actually perform an update (e.g. re-run search strategy, data- extraction, etc.) or retraction (e.g. removal from webpage, removal from portfolio, etc.).
Management options	A predefined option to handle an SR (conclusion) or CPG (recommendation) when the target condition is present or absent, depending on the specific stage, severity, status, or circumstance.
Outcome level	The level of measurement in a priority-setting assessment (see ordinal outcome and ranked continuous outcome).
Ordinal outcome	When the outcome of a priority-setting assessment has ordinal categories. Individual systematic reviews or clinical practice guidelines are designated by a priority-setting assessment to one of the ordinal categories indicating the level of priority.
POMBYTT framework	The Portfolio Maintenance by Test-Treatment (POMBYTT) framework is a framework based on a diagnostic test-treatment pathway (i.e. diagnosis, staging, management, and monitoring) to aid the design and tailoring of a maintenance strategy.[12]
Priority indicator	An indicator that signals the degree of priority of a systematic review or clinical practice guideline on a single factor.
Priority-setting assessment	A process to assess the degree of priority of (individual) systematic reviews or clinical practice guidelines. Priority-setting assessments may use priority indicators to assess the perceived priority.
Ranked continuous outcome	The outcome level of a priority-setting assessment that ranks (individual) systematic reviews or clinical practice guidelines on their priority. The resulting list shows individually ranked systematic reviews or clinical practice guidelines ordered from highest to lowest priority.
Role	The role a priority-setting can take in a maintenance strategy. Different roles have different performances in the maintenance strategy. Both the triage and add-on roles are described.
Sentinel	The characteristic of a priority-setting assessment which challenges all systematic reviews or clinical practice guidelines to demonstrate a certain degree of priority before allowing them to continue. The level or degree of priority needed to demonstrate is defined by a cut-off (see cut-off or limit of capacity). Systematic reviews or clinical practice guidelines not demonstrating that predefined level or degree of priority are deferred.
Triage	Triage is one of the roles in a priority-setting assessment. In this role, the priority-setting assessment is placed before diagnosis in the POMBYTT framework. Triage allows priority systematic reviews or clinical practice guidelines to continue within the maintenance strategy while lower-priority systematic reviews and guidelines are deferred.

and probability, organizational decisions, outdatedness of conclusions or recommendations, patient burden or inclusion, users, and other. Brief descriptions of priority-setting assessments and their outcome level are described in Table A2 (Additional file 1). Both ordinal and ranked continuous outcome levels were observed in the included literature. Six of the included references in the literature review reported an ordinal outcome of their priority-setting.[3, 15-19] Three references used a ranked priority-setting outcome.[2, 20, 21]

3.2 PRIORITY-SETTING ASSESSMENT ROLES IN THE POMBYTT FRAMEWORK

Where the POMBYTT framework (Figure A2 in Additional file 1) already used diagnostic and staging tests,[12] priority-setting assessments can be regarded as new tests measuring the degree of priority to be added to the framework. Like tests in diagnostic test-treatment strategies, priority-setting assessments can be placed at different positions in the POMBYTT framework defining their role. There are different downstream consequences for the SR or CPG after the priority-setting assessment depending on whether there was sufficient priority. These downstream consequences should be defined so that it is clear what happens to items with (in)sufficient priority in the maintenance strategy. Although re-assessment is defined for items with not enough priority in Figure 1 (red pathway), the downstream consequence could be described differently according to the organization's need in the maintenance strategy. For example, an alternative downstream consequence for insufficient priority in Figure 1B (red pathway) could be '*perform management actions as soon as capacity is freed*'.

3.2.1 TRIAGE ROLE

The priority-setting assessment in a triage role is placed before existing tests in the framework (i.e. diagnostic and staging tests), resulting in a triage before the target condition (e.g. outdatedness) is detected. The priority-setting assessment becomes the first test before any SR or CPG advances to further testing. Figure 1A schematically shows the position of a triage priority-setting assessment, while Figure 2 shows its position in the POMBYTT framework. We provide examples of indicators which could be used for triage priority-setting assessments in Table 2.

3.2.2 ADD-ON ROLE

The priority-setting assessment in an add-on role is placed after the existing tests (i.e. diagnosis and staging tests) and the decision for a specific management option in the framework. However, it is positioned before actually performing management actions associated with the selected management option. Thus, such assessment is only performed in the group of SRs or CPGs for which the management option was selected. Figure 1B schematically shows the position of an add-on priority-setting assessment. See Figure 2 for the positioning of an add-on priority assessment in the POMBYTT framework. Although the Figure 2 shows priority-setting after the decision to update was made, add-on priority setting can be used after the selection of any other management option when desirable. Table 3 shows examples of indicators which could be used in add-on priority-setting assessments.

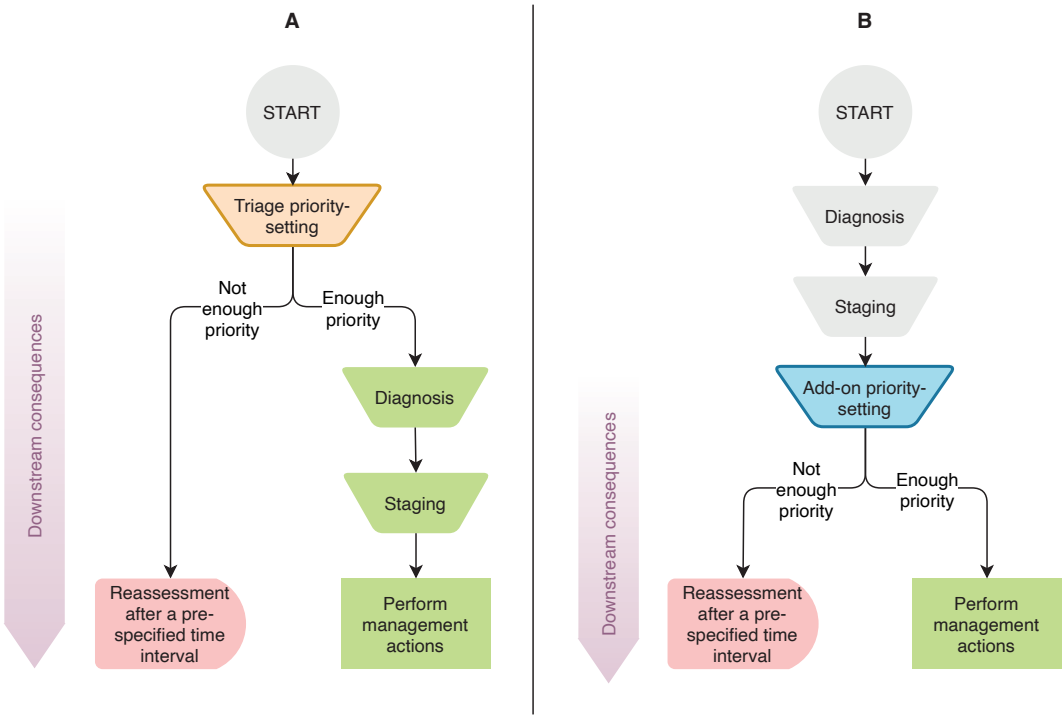
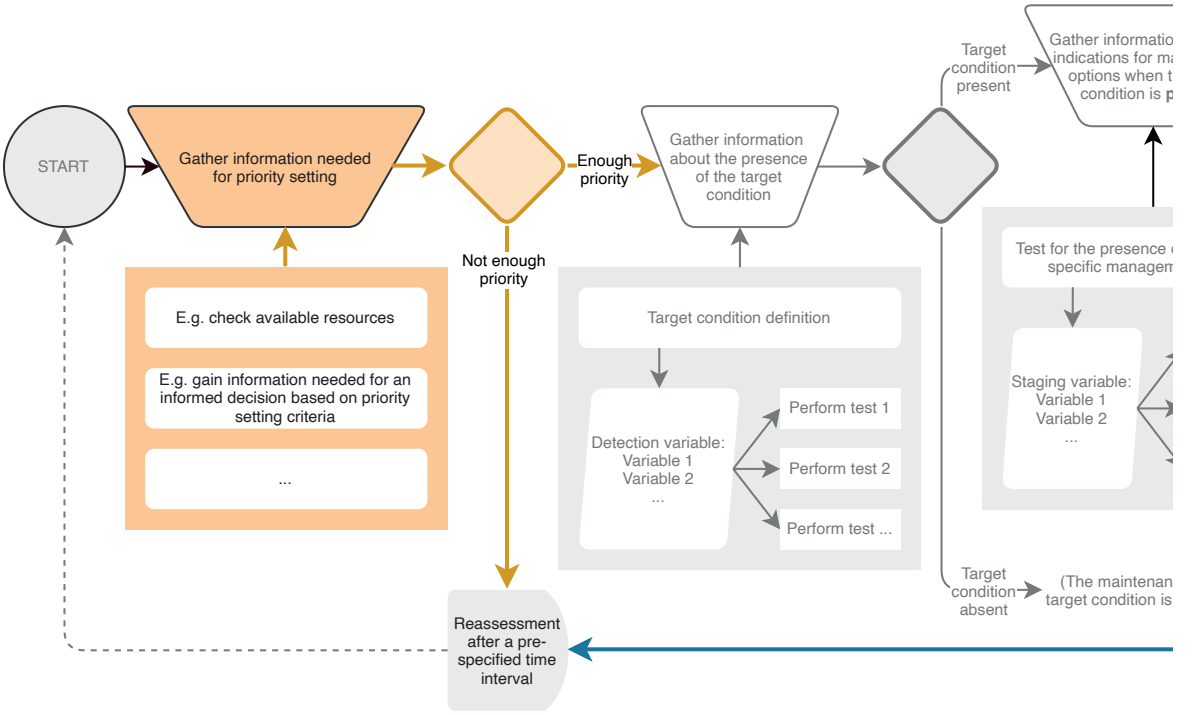


Figure 1. Schematic representation of priority-setting assessment roles. *The priority-setting in a triage role (A) is depicted in orange, while the priority-setting in the add-on role (B) is depicted in blue. Early triage priority-setting determines which priority systematic review or clinical practice guideline enters the strategy. Add-on priority-setting determines which priority systematic review or clinical practice guideline receives management actions. The priority-setting assessment's definitive outcome results in different downstream consequences for the SRs or CPGs (purple arrows), depending on whether the SRs or CPGs have (in)sufficient priority (green vs. red pathways in both A and B).*

3.3 PRIORITY-SETTING OUTCOME LEVELS

3.3.1 ORDINAL OUTCOME LEVEL

A priority-setting assessment with an ordinal outcome can have two or more levels of priority (see Figure 3A and 3B). The assessment measures priority of individual SRs or CPGs to assign them to the appropriate level of priority (e.g. low, medium, high, very high). The level of priority can only be distinguished between the groups of SRs or CPGs in each level due to the ordinal outcome. For example, a 'very high'-level has a higher priority than those in the 'high', 'medium', and 'low'-levels. However, it can be difficult to discern the degree of priority of individual SRs or CPGs within one ordinal level. For example, it might not be known which individual SR or CPG has a higher or lower degree of priority compared to other individual SRs or CPGs within the 'very high'-level. A cut-off can be used to discriminate between those groups of SRs and CPGs having enough priority to continue and those with insufficient priority (Figure 3A and 3B).


START
Gather information needed for priority setting
E.g. check available resources
E.g. gain information needed for an informed decision based on priority setting criteria
...
Enough priority
Not enough priority
Gather information about the presence of the target condition
Target condition definition
Detection variable:
Variable 1
Variable 2
...
Perform test 1
Perform test 2
Perform test ...
Target condition present
Gather informatio
indications for ma
options when t
condition is p
Test for the presence
specific managem
Staging variable:
Variable 1
Variable 2
...
Target condition absent
(The maintenan
target condition is
Reassessment after a pre-specified time interval

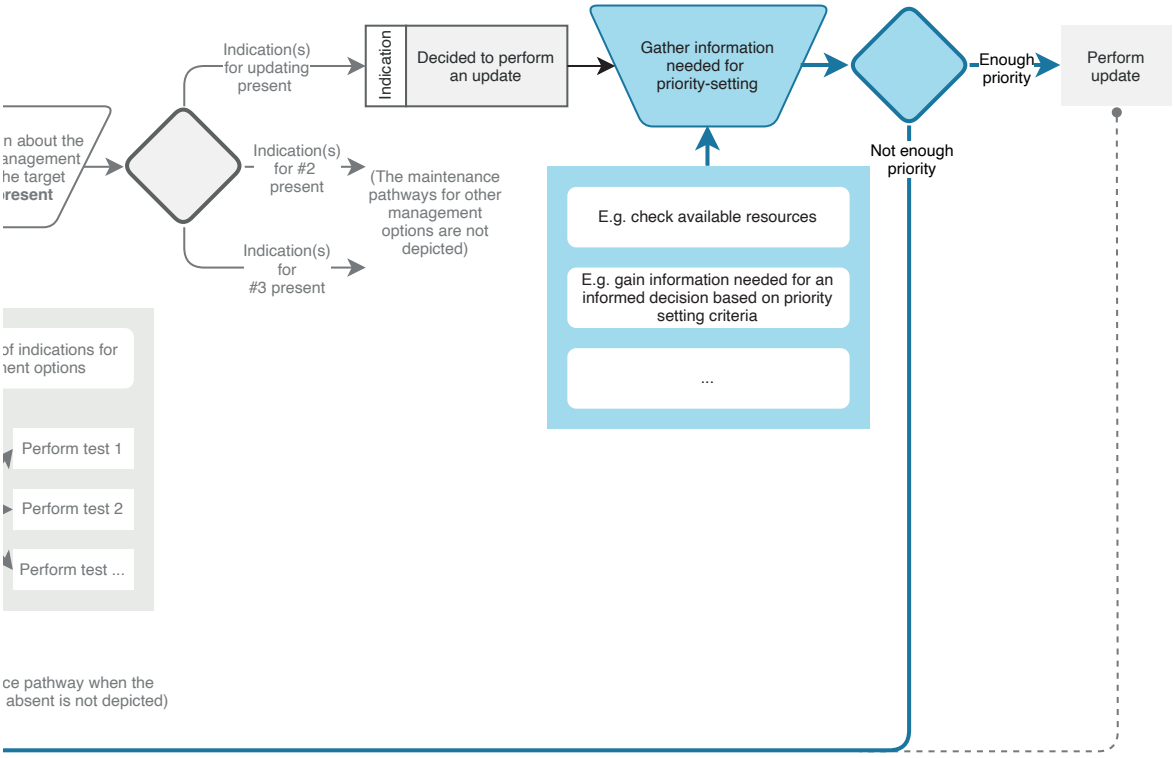


Figure 2. Priority-setting assessments in the Portfolio Maintenance by Test-Treatment framework. *The figure shows the position priority-setting in the triage role (orange) and in the add-on role (blue). Add-on priority-setting is placed after the decision for a specific management option is made but before the associated management actions are performed. Add-on priority-setting was only depicted for one management option in this figure, however add-on priority setting can be performed after the decision for any management option.*

Table 2 – Examples of priority indicators which could be used in a priority-setting assessment in the triage role

Domain	Indicator	Reference	Described for:
Access	Does the document have an Impact on access to care?	3	CPG
	Impact on access to health care	2	CPG
Clinical practice	Controversial areas	22	CPG
	Will a newly developed guideline have a meaningful impact on clinical decision-making? †	14	CPG
	Will a newly developed guideline have a meaningful impact on clinical outcomes? †	14	CPG
	Potential for a recommendation to affect clinical practice (based on existing controversy or the belief that a gap exists between evidence and practice)	19	CPG
	Are there substantial variations in the diagnosis and/or treatment of the health problem?	14	CPG
	Resolving areas of conflict within field	23	CPG
	The recommendation is a priority for decision-making	11	CPG
New evidence	Availability of new relevant evidence	2	CPG
	New evidence	22	CPG
	Expected volume of new evidence	2	CPG
	Present emergence of new evidence	23	CPG
	There are primary studies for inclusion in the review	24	SR
	Are there sufficient research findings available upon which to base a clinical practice guideline?	19	CPG
	There is likely to be new evidence	11	CPG
Organizational decisions	Will a newly developed guideline be of significant benefit to the organization's members? †	14	CPG
	Need for a balanced portfolio of topics	19	CPG
	Are potential new topics, nominated for guideline development, more important to the organization's membership? The Guideline committee will evaluate the value of updating a published guideline against potential new topics for guideline development	14	CPG
Patient burden or inclusion	The author team is agreeable to formally including consumers and/or other stakeholders in the review planning, conduct and/or dissemination	24	SR
Other	Are there existing guidelines on this topic?	14	CPG
	Context relevance of the clinical question	2	CPG
	Methodological applicability of the clinical question	2	CPG
	Development of new clinical questions	2	CPG

†From the described procedures it was deduced that updates were probably also prioritized according to this signal.
CPG: Clinical practice guideline
SR: Systematic review

Table 3 – Examples of priority indicators which might be used in a priority-setting assessment in the add-on role

Domain	Indicator	Reference	Described for:
Clinical practice	Will a newly developed guideline reduce practice variation?	[14]	CPG
New evidence	New evidence (e.g. new studies or new analyses of previous data) that had the potential to change the prior recommendation	[19]	CPG
	Is there known evidence that has been published since this document's last literature search that would result in significant changes to the recommendations?	[3]	CPG
	There is a reasonable chance that the existing recommendation changes with the emergence of new evidence	[11]	CPG
Prediction and probability	The likelihood of the effect size of the updated meta-analysis to lay inside, outside, or across prespecified limits representing clinical equivalence	[20]	SR
	Estimated probability of conclusions changing after the addition of new studies to an existing meta-analysis, by: estimated p = invlogit(0.1207 + 0.4101 x weight ratio + 0.1836 x number of new trials)	[21]	SR
	The ratio of the sum of weights allotted to the predicted new and existing studies in the updated meta-analysis	[20]	SR
	The probability of producing statistically significant results when adding further studies (in which the results are consistent with those that already exist) to an existing meta-analysis	[20]	SR
	The number of additional participants (on average) from new evidence relative to the predicted number of participants required to obtain significance	[20]	SR
	The ratio of the standard error of the predicted new estimate of effect to the existing effect	[20]	SR
Organizational decisions	Is there substantial public or political demand for this practice guideline?	[14]	CPG
Outdatedness of conclusions or recommendations	Impact of outdated recommendations on safety	[2]	CPG
	How out of date is the part of the systematic review?	[16-18]	SR
	How out of date are the conclusion considering potential changes in practice or therapy preference	[15]	SR
	How out of date are the conclusion considering the magnitude/direction of changes in estimates	[15]	SR
	How out of date are the conclusion considering the availability of new treatment	[15]	SR
	How out of date are the conclusion considering safety issues including withdrawn from the market drugs/ black box warning	[15]	SR
	How many conclusions of the systematic review are up-to-date, possibly out of date, or certainly out of date?	[15]	SR
	How much of the systematic review is possibly, probably or certainly out of date?	[15]	SR

Patient burden or inclusion	Does the health problem in question carry a high individual or population burden of morbidity, mortality, or disability?	14	CPG
	Public health importance (i.e. burden of suffering and expected effectiveness of the preventive service to reduce that burden)	19	CPG
Users	According to users' need for new and updated guidelines	25	CPG
	User's interest	2	CPG
Other	Does the health problem and/or its diagnosis and/or treatment carry a high unit or aggregate cost?	14	CPG
	Item scores and priority scores†	2	CPG
	The review can be commenced in a timely manner	24	SR

†Scores calculated with a priority-setting tool
CPG: Clinical practice guideline
SR: Systematic review

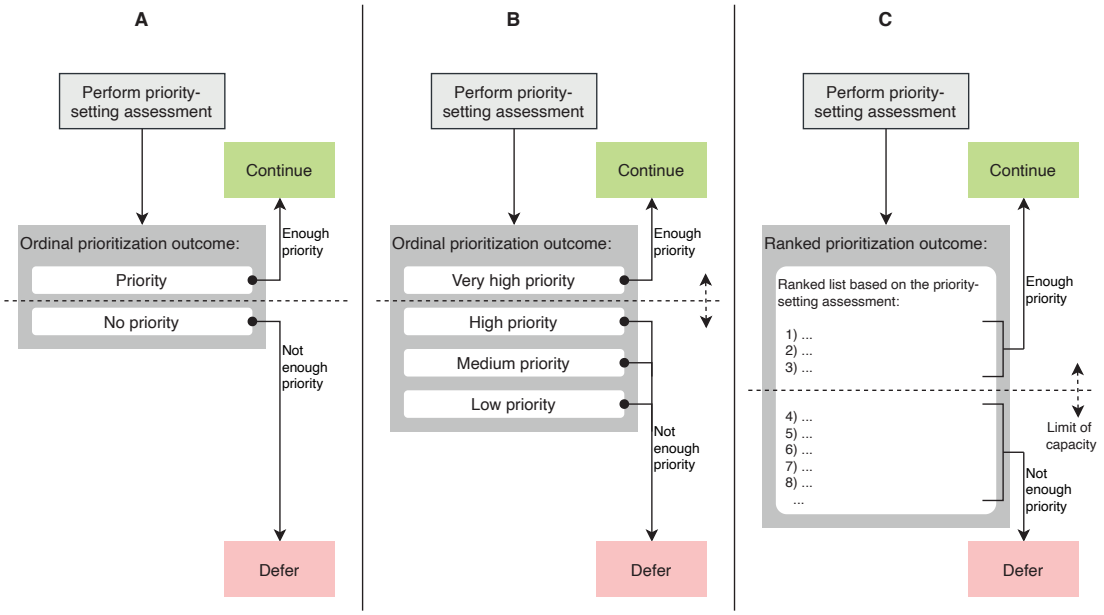


Figure 3. Schematic representation of priority-setting outcome levels. *The figure shows priority-setting assessments with and ordinal outcome level having two (A) or more (B) ordinal levels of priority. Furthermore, a ranked continuous priority-setting assessment is showed (C). The dashed line represents the cut-off, determining which items have (in) sufficient priority. The cut-off can move (dashed arrow in B and C) to expand or limit the definition of (in)sufficient priority. In a ranked continuous outcome (C) the cut-off can be aligned to the available resources, making it a limit of capacity (dashed arrow in C). The sentinel characteristic of priority-setting can be observed here, where items with high enough priority continue (green box) and items with not enough priority are deferred (red box).*

3.3.2 RANKED CONTINUOUS OUTCOME LEVEL

The priority-setting assessment with a ranked continuous outcome will result in a ranked list of SRs or CPGs. The degree of priority of individual SRs or CPGs is measured and compared to the degree of priority of other individual SRs or CPGs. Any arbitrary cut-off can be used to select the top-ranked SRs or CPGs with the highest degree of priority. However, with resource efficiency considered, a limit of capacity can be used as a cut-off to select the top ranked SRs or CPGs (see Figure 3C). The limit of capacity translates the available resources an organization can or is willing to invest into a specific number of SRs or CPGs (i.e the capacity). Here, we view resources in a broad sense (e.g. time, budget, available personnel).

3.4 PERFORMANCE OF PRIORITY-SETTING ASSESSMENT IN MAINTENANCE STRATEGIES

Priority-setting in a maintenance strategy can be considered as a sentinel (Figure 3). All SRs or CPGs entering the priority-setting assessment are challenged to demonstrate a certain level or degree of priority to pass through. How a priority-setting assessment performs in a maintenance strategy depends on its role and its outcome level.

In a triage role, the priority-setting assessment performs like a sentinel controlling the inflow of SRs or CPGs in the maintenance strategy, where only those SRs or CPGs with enough priority may continue for further testing in the maintenance strategy. In practice, it is a priority-check to test whether SRs or CPGs have a certain level of priority before resources are invested for further testing. Priority-setting in the add-on role will identify those SRs or CPGs which have a certain priority to receive the management option. The add-on priority-setting assessment performs as a sentinel controlling the outflow after testing in the framework, where only those SRs and CPGs with enough priority receive the management actions associated with the selected management option. In other words, it is a priority-check to test which SRs or CPGs have enough priority to spend the limited resources for the selected management option on.

A priority-setting assessment using an ordinal outcome does not limit the SRs or CPGs to a specific or prespecified amount continuing in the strategy. Instead, the assessment performs as a sentinel checking the perceived level of priority. Each SR or CPG demonstrating a certain level (e.g. ‘very high priority’) or degree of priority continues, while SRs and CPGs with a lower priority get deferred (see Figure 3). The priority-setting assessment using a ranked continuous outcome informs how SRs or CPGs are ranked based on their perceived priority. Using a cut-off, or a limit of capacity, the priority-setting assessment performs as a sentinel letting those top-ranked SRs or CPGs pass through. This limits the number of SRs or CPGs passing through to a specific amount. A ranked continuous priority-setting outcome using a cut-off will always let at least one SR or CPG continue, since the smallest cut-off lies between rank one and rank two. An ordinal priority-setting outcome may theoretically not let any SR or CPG continue when they all demonstrated a lower level of priority than the cut-off.

Four distinct theoretical performances of a priority-setting assessment in a maintenance strategy can be described based on the combination of the role and outcome level. Table 4 summarizes the performances for each combination of role and outcome level in a two-by-

two table.

3.5 ADDING PRIORITY-SETTING ASSESSMENTS TO A MAINTENANCE STRATEGY

Additional file 1 provides a blank description table for priority-setting assessments in maintenance strategies (see Table A3). The role, outcome level, and cut-off can be defined in such table. The table furthermore allows to describe the downstream consequences of the priority-setting's definitive outcome. The downstream consequences should describe what happens to the SR or CPG in the maintenance strategy when it had (not) enough priority. This could depend on the role of the priority-setting assessment has in the maintenance strategy. For illustrative purposes, a previously designed hypothetical 'living' example of an updating strategy was extended with two priority-setting assessments (see Table A4 and Figure A3 in Additional file 1).

Table 4 –Description of the theoretical performance of a priority-setting assessment in a maintenance strategy depending on its role and outcome level.

	Triage	***Add-on***
Ordinal outcome	Identifies a group of systematic reviews or clinical practice guidelines that has a higher level of priority to continue in the maintenance strategy compared to other group(s) of systematic reviews or clinical practice guidelines. Theoretically, none of the assessed systematic reviews or clinical practice guidelines may demonstrate the level of priority needed to continue in the maintenance strategy.	Identifies a group of systematic reviews or clinical practice guidelines that has a higher level of priority to continue and receive management action associated with the selected management option compared to other group(s) of systematic reviews or clinical practice guidelines. Theoretically, none of the assessed systematic reviews or clinical practice guidelines may demonstrate the level of priority needed to continue in the maintenance strategy.
Ranked continuous outcome	Identifies which individual systematic review or clinical practice guideline has a higher or lower degree of priority compared to other individual systematic reviews or clinical practice guidelines. When using a cut-off only a specific number of top-ranked systematic reviews or clinical practice guidelines continue in the maintenance strategy (at least one).	Identifies which individual systematic review or clinical practice guideline has a higher or lower degree of priority compared to other individual systematic reviews or clinical practice guidelines. When using a cut-off only a specific number of top-ranked systematic reviews or clinical practice guidelines will continue (at least one) and receive the management actions that are associated with the selected management option.

4. DISCUSSION

4.1 THE FRAMEWORK EXTENSION IN CONTEXT

In an ideal world all SRs and CPGs are immediately updates as soon as new data becomes available. However, the sheer number of available SRs and CPGs, the rate in which new data becomes available, and the necessary (human) resources combined makes it arguably impossible to keep up. Choices must be made, and priority-setting can aid when resources are limited. Rather than designing a priority-setting assessment, the extension of the framework describes how priority-setting assessments could perform in a maintenance strategy of an SR or CPG portfolio, depending on their role and outcome level. We have used the roles described in diagnostic test accuracy methodology[13] for the extension of the framework. The outcome level and cut-off in priority-setting assessments determine how the sentinel characteristic performs when priority-setting assessments are added to a maintenance strategy. The framework supports the design of maintenance strategies under resource constraints. The organization's intended purpose of priority-setting in a resource constrained maintenance strategy could then be aligned with a specific performance to select the right outcome level and role for their priority-setting assessment.

The resource costs of the maintenance strategy and its management options are important to consider. For example, a living CPG recommendations strategy may be resource intensive for each recommendation in the portfolio as continuous monitoring and updating are costly. It is proposed to prioritize recommendations for assigning a living status based on three indicators.[11] Recommendations are thus tested for their priority before entering the living maintenance strategy. The implications of the outcome level of such priority-setting assessment becomes important as there may be limited resources to perform the continuous monitoring. An ordinal outcome could result in more living recommendations requiring continuous monitoring and updating then the organization has resources available. A continuous ranked outcome with a limit of capacity could result in a specific number of living recommendations entering the living strategy. Although this may align with the available resources for monitoring, it could result in underutilizing the available resources for updating. Making the monitoring process less resource intensive could result in more recommendations being able to receive the living status. This, in turn, could allow for a better utilization of available resources for performing updates. The monitoring process might be made less resource intensive in the future by using artificial intelligence for literature monitoring, screening and selections. When this results in more updates than available resources allow, a second ranked continuous priority-setting assessment with a limit of capacity can be added to the strategy as an add-on.

In our literature review of priority indicators, we found 50 indicators which could be used in priority-setting assessments. During the literature review we encountered indicators in assessments concerning resource requirements in a broad sense.[2, 3, 14, 24, 26, 27] For example, *"Do the working groups responsible for this document have the resources available to write a full update of this document within the next year?"*[3] and *"Is a review team available".*[27] We did not classify these indicators as being priority indicators for the POMBYTT framework extension, although they may be considered as such by others. In essence, we believed

that limited resources itself is a reason to perform a priority-setting assessment rather than being informative for the level or degree of priority. An SR or CPG with no available author team could theoretically have the highest priority to receive management actions (e.g. performing an update). Such considerations may play a role in determining the priority-setting assessment's definitive outcome. In the UpPriority tool, for example, a definitive decision is made for which items are prioritized as a subsequent step after the items are ranked on their priority score.[1, 2] This decision is based on additional considerations besides these scores.[1, 2]

A previously published systematic review identified 76 priority indicators for updating.[6] There is only a small overlap of included studies between our literature review and that published systematic review. Differences in the search strategy and eligibility criteria could have caused a different inclusion of relevant studies. We also have excluded some studies[27, 28] which we perceived as being strategies assessing whether there is a need for updating rather than assessing priority. We considered the need for updating and the priority-setting to be distinct constructs. Furthermore, a second systematic review found 106 priority indicators to prioritize topics or questions for evidence syntheses[9] and a third systematic review previously found 118 priority indicators.[8] We excluded any priority indicator from our own literature review when it was not described in the context of SRs or CPGs, or when it was solely described for prioritizing new topics or questions. Nonetheless, the reported priority indicators in these systematic reviews could still be useful for priority-setting in a maintenance strategy.

5

4.2 CONSIDERATIONS FOR PRIORITY INDICATORS AND INTERPRETING ASSIGNED SCORES

Earlier frameworks of indicators have been published[8, 9] and an international Delphi panel reached consensus on a set of indicators for a prioritization tool.[2] However, different contexts such as variations in health care systems may require different priority indicators. For example, *"impact on access to health care"*[2] as an indicator might not be useful for priority-setting in healthcare systems where the system enables high accessibility by default. Furthermore, distinct organizational needs and policies might require different priority indicators. For example, local or national policies may require patient or other stakeholder perspectives in prioritization. Thus, sets of priority indicators may not necessarily be interchangeable between organizations. Reaching consensus among the organization and its stakeholders about the set of priority indicators used within the organization or for a specific context might currently be a good option.

It furthermore seems that priority indicators are regularly used in priority-setting assessments.[7] At least two different perspectives might be taken when interpreting scores assigned to such priority indicators. The first perspective assumes that there is a 'true' underlying value of priority as a characteristic of the SR or CPG being evaluated. Participants assigning scores are assessors evaluating the level of priority based on the priority indicators. Thus the effect of priority on the indicator is measured.[29] The assigned scores from individual participants in the priority-setting assessment then deviate from the 'true' value of priority as described in the classical test theory (i.e. *observed indicator value = true value ± error*[29]). The second perspective assumes that there is no underlying 'true' value of priority as a

characteristic of the SR or CPG. All participants in a priority-setting assessment may perceive the level of priority differently without being wrong or deviating from a 'true' value of priority as a measurable characteristic of the SR or CPG. The level or degree of priority of an SR or CPG is constructed from the assessors' perceived priority.

4.3 PRIORITY-SETTING ASSESSMENTS IN MAINTENANCE STRATEGIES

It is important to match the performance of a priority-setting assessment to the intended goal of priority-setting when designing a maintenance strategy. Different combinations of roles and outcome levels seem to result in different performances. A triage priority-setting assessment could be used when the goal is to limit SRs or CPGs entering the maintenance strategy by a specific number when using a ranked continuous outcome and a cut-off. However, if an organization wishes to only invest resources in SRs or CPGs with a certain level of priority regardless of the quantity, a triage priority-setting assessment with an ordinal outcome can be used. Priority-setting in the add-on role could be used when there are limited resources for the management actions associated with the selected management option by using a ranked continuous outcome with a limit of capacity. When only those SRs or CPGs with a certain level of priority should receive the management actions associated with the selected management option, an add-on priority-setting assessment with an ordinal outcome could be used. The ordinal outcome ensures that resources are spend on priority SRs or CPGs, however it does not limit the SRs or CPGs receiving the management actions associated with the selected management option to a specific number. The appropriate performance (and thus the appropriate role and outcome level) can probably more easily be selected when the purpose of the priority-setting in the maintenance strategy is made explicit when designing the strategy.

4.4 LIMITATIONS

The four described performances of priority-setting assessments are only theoretically in nature and were not tested in practice. Additional distinct performances may exist depending on other characteristics than the role and outcome level. Due to the theoretical nature of the description, priority-setting assessments might perform differently in practice than described or foreseen. Furthermore, we only selected literature in our review that reported a priority indicator for updating. The reported priority indicators were consequently described in an updating strategy context rather than a maintenance strategy context. It is unknown whether pathways leading to other management options should have the same or a different set of priority indicators when a priority-setting assessment is added.

4.5 FUTURE DIRECTIONS

Pilots using single or multiple priority-setting assessments with different roles and outcome levels should be performed and evaluated to provide some evidence about their performance in a maintenance strategy. Organizations have the flexibility to develop their own priority-setting assessments and incorporate them into their maintenance strategy using the POMBYTT framework. Additionally, organizations may also use published prioritization

tools, such as the UpPriority tool[1, 2] in triage or add-on priority-setting assessments when they fit in with the purpose of priority-setting in the maintenance strategy. Published tools can be adapted to accommodate different sets of priority indicators or modified processes relevant to the needs of the organization and stakeholders. Experiences with one or multiple priority-setting assessments having different roles and outcome levels in a maintenance strategy and their feasibility should be shared.

5. CONCLUSION

Priority-setting is necessary when limited resources prevent maintaining the entire portfolio of SRs or CPGs. The performance of a priority-setting assessments should be aligned with the goal of priority-setting in the maintenance strategy. This is even more important when the priority-setting process aims to optimize the utilization of the available resources. The described performances of a priority-setting assessment in the current POMBYTT framework extension provide guidance to the designers of a maintenance strategy regarding the appropriate role and outcome level for such assessments.

LIST OF ABBREVIATIONS

CPG: Clinical Practice Guideline
POMBYTT: Portfolio Maintenance by Test-Treatment
SR: Systematic Review

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

All data generated or analyzed during this study are included in this published article and its supplementary information file.

COMPETING INTERESTS

All authors have completed the ICMJE disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare:

M.S. Oerbekke, Msc: No conflicts of interest
R.G. Elbers, PhD: No conflicts of interest
M.J. Van der Laan, MSc, MD, PhD: No conflicts of interest
L. Hooft, PhD: No conflicts of interest

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

All author contributions are made transparent using the structured Contributor Role Taxonomy (CRediT) described at https://doi.org/10.1002/leap.1210.

MSO: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, Writing – review & editing.
RGE: Investigation, Writing – review & editing.
MJvdL: Conceptualization, Methodology, Writing – review & editing.
LH: Conceptualization, Methodology, Writing – review & editing.

ACKNOWLEDGEMENTS

None.

LITERATURE LIST

1. Sanabria AJ, Alonso-Coello P, McFarlane E, Niño de Guzman E, Roqué M, Martínez García L. The UpPriority tool supported prioritization processes for updating clinical guideline questions. J Clin Epidemiol. 2021;139:149-59.
2. Sanabria AJ, Pardo-Hernandez H, Ballesteros M, Canelo-Aybar C, McFarlane E, Niño de Guzman E, et al. The UpPriority tool was developed to guide the prioritization of clinical guideline questions for updating. J Clin Epidemiol. 2020;126:80-92.
3. Agbassi C, Messersmith H, McNair S, Brouwers M. Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. J Clin Epidemiol. 2014;67(12):1335-42.
4. Wiercioch W, Nieuwlaat R, Zhang Y, Alonso-Coello P, Dahm P, Iorio A, et al. New methods facilitated the process of prioritizing questions and health outcomes in guideline development. J Clin Epidemiol. 2022;143:91-104.
5. Zaror C, Deana NF, Espinoza-Espinoza G, Aravena-Rivas Y, Muñoz-Millán P, Pineda P, et al. Questions and health outcomes prioritization for the development of a COVID-19 dental clinical practice guideline: A case study. J Eval Clin Pract. 2022;28(3):404-10.
6. Martínez García L, Pardo-Hernandez H, Superchi C, Niño de Guzman E, Ballesteros M, Ibargoyen Roteta N, et al. Methodological systematic review identifies major limitations in prioritization processes for updating. J Clin Epidemiol. 2017;86:11-24.
7. El-Harakeh A, Lotfi T, Ahmad A, Morsi RZ, Fadlallah R, Bou-Karroum L, Akl EA. The implementation of prioritization exercises in the development and update of health practice guidelines: A scoping review. PLoS One. 2020;15(3):e0229249.
8. El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. BMC Health Serv Res. 2019;19(1):692.
9. Fadlallah R, El-Harakeh A, Bou-Karroum L, Lotfi T, El-Jardali F, Hishi L, Akl EA. A common framework of steps and criteria for prioritizing topics for evidence syntheses: a systematic review. J Clin Epidemiol. 2020;120:67-85.
10. Becker M, Jaschinski T, Eikermann M, Mathes T, Bühn S, Koppert W, et al. A systematic decision-making process on the need for updating clinical practice guidelines proved to be feasible in a pilot study. J Clin Epidemiol. 2018;96:101-9.
11. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol. 2017;91:47-53.
12. Oerbekke MS, Elbers RG, van der Laan MJ, Hooft L. Designing tailored maintenance strategies for systematic reviews and clinical practice guidelines using the Portfolio Maintenance by Test-Treatment (POMBYTT) framework. BMC Med Res Methodol. 2024;24(1):29.
13. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: asssing new tests against existing diagnostic pathways. Bmj. 2006;332(7549):1089-92.
14. American Academy of Orthopaedic Surgeons Clinical Practice Guideline Unit. Guideline Process Procedure for Updating Existing Evidence-Based Clinical

Practice Guidelines. American Academy of Orthopaedic Surgeons; 2011.
15. Ahmadzai N, Newberry SJ, Maglione MA, Tsertsvadze A, Ansari MT, Hempel S, et al. A surveillance system to assess the need for updating systematic reviews. Syst Rev. 2013;2:104.
16. Newberry SJ, Ahmadzai N, Motala A, Tsertsvadze A, Maglione M, Ansari MT, et al. AHRQ Methods for Effective Health Care. Surveillance and Identification of Signals for Updating Systematic Reviews: Implementation and Early Experience. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
17. Shekelle PG, Motala A, Johnsen B. AHRQ Methods for Effective Health Care. Assessment of a Method to Detect Signals for Updating Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
18. Shekelle PG, Motala A, Johnsen B, Newberry SJ. Assessment of a method to detect signals for updating systematic reviews. Syst Rev. 2014;3:13.
19. US Preventive Services Task Force. US Preventive Services Task Force Procedure Manual. 2015.
20. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. J Clin Epidemiol. 2009;62(3):241-51.
21. Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. Bmj. 2013;347:f7191.
22. World Health Organization. WHO handbook for guideline development - 2nd edition. Geneva, Switzerland: WHO Press; 2014.
23. Jiang ZM, Zhan SY, Zuo L, Jia XW, Fang H, Li XX, et al. Guideline for [Clinical Guidelines Constitution/Amendment] in China. Int J Qual Health Care. 2019;31(7):568-74.
24. Synnot AJ, Tong A, Bragge P, Lowe D, Nunn JS, O'Sullivan M, et al. Selecting, refining and identifying priority Cochrane Reviews in health communication and participation in partnership with consumers and other stakeholders. Health Res Policy Syst. 2019;17(1):45.
25. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual (April 2017 update). 2014.
26. U.S. Preventive Services Task Force. US Preventive Services Task Force Procedure Manual. 2015.
27. Welsh E, Stovold E, Karner C, Cates C. Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach. J Clin Epidemiol. 2015;68(3):341-6.
28. Soll RF. Updating reviews: the experience of the Cochrane Neonatal Review Group. Paediatr Perinat Epidemiol. 2008;22 Suppl 1:29-32.
29. DeVellis RF. Classical test theory. Med Care. 2006;44(11 Suppl 3):S50-9.

CHAPTER 5
Additional files

ADDITIONAL FILE 1

A1.1 REVIEW METHODS

A formal literature review protocol was not prepared a priori and the literature review was not registered.

A1.1.1 SEARCH STRATEGY

Preliminary searches indicated that 'prioritizing' and 'need for updating' might be used interchangeably in the literature for what we considered to be separate constructs. Therefore, we combined 'prioritizing' with 'need for updating' and constructed the following search string: ((need for updating) OR (prioritizing)) AND (methods) AND (systematic review OR guideline). On the 10th of October 2018 we searched MEDLINE (through PubMed) and on the 29th of April 2020 we updated the search. Furthermore, PubMed Health was searched and the results were limited by the 'methods resources' filter through the search engine's interface. We could not update the PubMed Health search since the database was discontinued on 31 October 2018.[1] Appendix A1.1.1.1 and A1.1.1.2 show the complete search strings. Handbooks identified in a previously published systematic review (SR) were obtained through their respective organization's website when available.[2] Reference lists from the included literature were hand searched for additional journal articles, reports, and handbooks

A1.1.1.1 PUBMED SEARCH STRING

((need[tiab] OR needs[tiab] OR needed[tiab] OR signal[tiab] OR signals[tiab] OR require[-tiab] OR requires[tiab] OR required[tiab] OR detect[tiab] OR when[tiab]) AND (update[tiab] OR updates[tiab] OR updating[tiab] OR updated[tiab] OR "out of date"[tiab])) OR (prioritize[-tiab] OR prioritise[tiab] OR prioritization[tiab] OR prioritisation[tiab] OR priority[tiab] OR priorities[tiab] OR prioritized[tiab] OR prioritised[tiab] OR prioritizing[tiab] OR prioritising[tiab]) AND ("Models, Theoretical"[Mesh] OR method[tiab] OR methods[tiab] OR methodology[tiab] OR methodologies[tiab] OR process[tiab] OR processes[tiab] OR surveillance[tiab] OR strategy[tiab] OR strategies[tiab] OR system[tiab] OR systems[tiab] OR approach[tiab] OR approaches[tiab] OR criteria[tiab] OR checklist[tiab] OR tool[tiab] OR content[tiab]) AND ("Review literature as topic"[mesh] OR "meta-analysis as topic"[mesh] OR "Practice guidelines as topic"[mesh] OR "guidelines as topic"[mesh] OR "Time factors"[mesh])

A1.1.1.2 PUBMED HEALTH SEARCH STRING

FILTER: METHODS RESOURCES

((need[tiab] OR needs[tiab] OR needed[tiab] OR signal[tiab] OR signals[tiab] OR require[-tiab] OR requires[tiab] OR required[tiab] OR detect[tiab] OR when[tiab]) AND (update[tiab] OR updates[tiab] OR updating[tiab] OR updated[tiab] OR "out of date"[tiab])) OR (prioritize[-tiab] OR prioritise[tiab] OR prioritization[tiab] OR prioritisation[tiab] OR priority[tiab] OR priorities[tiab] OR prioritized[tiab] OR prioritised[tiab] OR prioritizing[tiab] OR prioritising[tiab])

A1.2 ELIGIBILITY AND LITERATURE SELECTION

We included a study, report, or handbook which described at least one indicator that informed the prioritization of SRs or (sections of) clinical practice guidelines (CPGs) for updating. We excluded literature when a process was described that solely identified whether SRs or (sections of) CPGs were in need for updating without prioritization. Literature was also excluded when prioritization was described in other contexts than updating SRs or CPGs (e.g. research priorities or de novo topics only). We furthermore excluded literature when it was a conference abstract, oral or poster presentation, (updated) CPG or SR without original relevant data, or was a report on obvious non-relevant topics (e.g. intervention, diagnosis, prognosis, animals, plants, cost-effectiveness, health technology assessment). One author (MSO) removed all obvious non-relevant results yielded by the search strategy. If there was any doubt, the reference was advanced to the title and abstract screening phase. Two authors (MSO, RGE) performed the title and abstract screening and were blinded for each other's decisions. We also advanced references to the full-text selection phase when in doubt or disagreement. The remaining full-text journal articles, reports, and handbooks were read full-text by two authors (MSO, RGE) who were blinded for each other's selection. A third author (LH) was consulted to resolve any disagreements after discussion.

A1.3 DATA EXTRACTION AND DATA HANDLING

Indicators informing about the priority for updating a SR or CPG and some general characteristics were extracted by one author (MSO) using a standardized data-extraction form. We generalized indicators that included specific names, fields, or professions. For example, when the indicator was "Is performing this guideline feasible with current AAOS resources?",[3] it was generalized to "Is performing this guideline feasible with the organization's current resources?". Indicators presented clustered, or in one sentence, were separated when we thought it was reasonable to assume that they could be independent of each other.

A1.4 DATA ANALYSIS

We initially used NVivo for Windows (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12) for qualitative analyses to identify some common domains among the extracted indicators. Indicators were assigned only one label during the quantitative analyses to prevent a duplicate representation under multiple domains. Qualitative coding was performed by one author (MSO) and the resulting emerging domains and their contents were discussed with another author (LH). Changes resulting from these discussions were realized without the use of NVivo. We furthermore placed any additional indicators resulting from the updated search strategy in the existing domains without a re-analysis in NVivo.

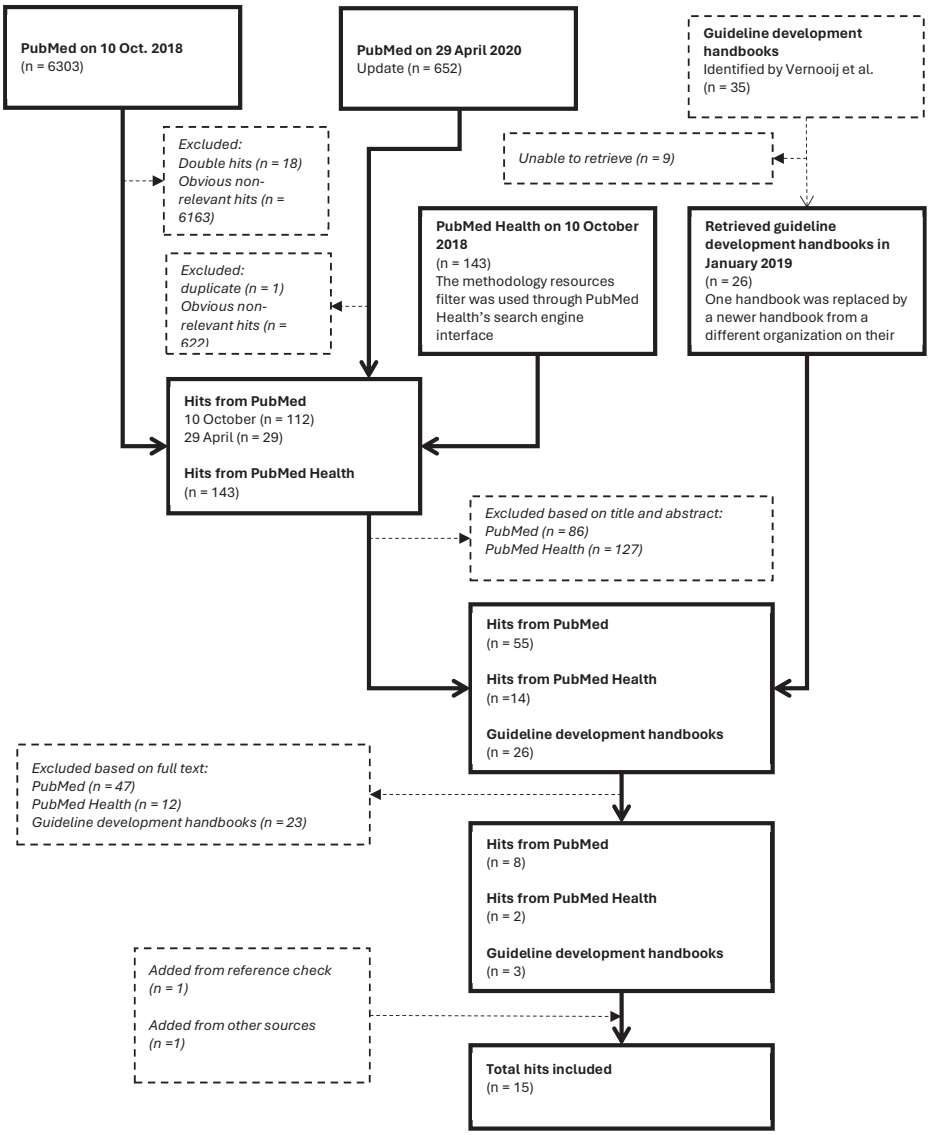


Figure A1. Flow diagram of the study selection. *Handbooks were identified through a previously published systematic review by Vernooij et al.*[2]

LITERATURE LIST

1. NCBI Insights. PubMed Health to be discontinued October 31, 2018; content will continue to be available at NLM. 2018. Available from: https://ncbiinsights.ncbi.nlm.nih.gov/2018/07/31/pubmed-health-discontinued-october-2018-nlm/
2. Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. Implement Sci. 2014;9:3.
3. American Academy of Orthopaedic Surgeons Clinical Practice Guideline Unit. Guideline Process Procedure for Updating Existing Evidence-Based Clinical Practice Guidelines. American Academy of Orthopaedic Surgeons; 2011.

Table A1 – Reasons for exclusion

Author or organization	Year	Title	Reason for exclusion
Agencia d'avaluacio de technologia i recerca mediques		Guies de pràctica clínica.	Could not be found online
AHCPR	1994	Process for determining need for updates of clinical practice guidelines--AHCPR	No priority indicators were found
Alonso-Coello	2011	The updating of clinical practice guidelines: insights from an international survey	No priority indicators were found
American Academy of Otolaryngology head and neck surg	2013	Clinical Practice Guideline Development Manual, Third Edition: a quality-driven approach for translating evidence into action.	No priority indicators were found
American College of Chest Physicians		Evidence-based Guideline Development Process	No priority indicators were found
American College of Occupational and environmental medicine	2017	Methodology for ACOEM's Occupational Medicine Practice Guidelines – 2017 Revision	No priority indicators were found
American College of Physicians	2010	The development of clinical practice guidelines and guidance statements of the American College of Physicians: summary of methods.	No priority indicators were found
American heart association / American college of cardiology foundation	2010	Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines.	No priority indicators were found
American society of clinical oncology		American Society of Clinical Oncology Guideline Procedures Manual.	Could not be found online
American urological association	2015	Overview: Standard Operating Procedures.	No priority indicators were found
Arzneimittelkommission der deutschen ärtzeschaft	2011	Leitfaden für die Erstellung von Therapieempfehlungen.	No priority indicators were found
Ärztliche Zentralstelle Qualitätssicherung		National Disease Management Guidelines: Method Report.	Could not be found online
Atkins	2012	Priority setting in guideline development: article 2 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report	No priority indicators were found
Barrowman	2003	Identifying null meta-analyses that are ripe for updating	No priority indicators were found
Bashir	2018	Time-to-update of systematic reviews relative to the availability of new evidence	No priority indicators were found

Bastian	2011	Choosing health technology assessment and systematic review topics: the development of priority-setting criteria for patients' and consumers' interests	Seems to cover priority-setting of new topics
Battista	1995	Setting priorities and selecting topics for clinical practice guidelines	Seems to cover priority-setting of new topics
Becker	2014	Partial updating of clinical practice guidelines often makes more sense than full updating: a systematic review on methods and the development of an updating procedure	No priority indicators were found
Bero	2013	The Cochrane Collaboration review prioritization projects show that a variety of approaches successfully identify high-priority topics	Seems to cover priority-setting of new topics
Bundersärtzekammer		National Disease Management Guidelines.	Could not be found online
Burgers	2012	Adaptation, evaluation, and updating of guidelines: article 14 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report	No priority indicators were found
Canadian Medical Association	2007	Handbook on Clinical Practice Guidelines.	No priority indicators were found
Canadian thoracic society	2007	Canadian Thoracic Society: Presenting a new process for clinical practice guideline production.	No priority indicators were found
Caring for Australasians with Renal Impairment	2015	A Guide for Writers.	No priority indicators were found
Centrul Naţional de Studii Medicina Familiei		Metodologie elaborarii ghidului de practica.	No priority indicators were found
Chung	2012	Two methods provide similar signals for the need to update systematic reviews	No priority indicators were found
D. Pieper	2019	[Increasing the efficiency of guideline production: a narrative review]	No original data
Dalal	2012	A Pilot Study Using Machine Learning and Domain Knowledge To Facilitate Comparative Effectiveness Review Updating [Internet]	No priority indicators were found
Domus Medical Flemish College of General Practitioners34		Algemeen Stramien voor de Ontwikkeling van Aanbevelingen van Goede Medische Praktijkvoering.	Could not be found online
Doyle	2005	Global priority setting for Cochrane systematic reviews of health promotion and public health research	No priority indicators were found, might concern new topics
Drug Commission of the German Medical Association	2006	Handbuch zur Entwicklung regionaler Leitlinien.	No priority indicators were found

Duodecim Finnish Medical Society		Submitted NHS Evidence Accreditation Application.	Could not be found online
Duodecim Medical Publications		Preface: What is Evidence-Based Medicine Guidelines.	Could not be found online
El-Harakeh	2019	Prioritization approaches in the development of health practice guidelines: a systematic review	Seems to cover priority-setting of new topics
European Region Of The World Confederation For Physical Therapy		Framework for Clinical Guideline Development in Physiotherapy.	Could not be found online
European Society of Gastrointestinal Endoscopy	2012	European Society of Gastrointestinal Endoscopy (ESGE) Guideline Development Policy	No priority indicators were found
French	2005	Investing in updating: how do conclusions change when Cochrane systematic reviews are updated?	No priority indicators were found
Gartlehner	2004	Assessing the need to update prevention guidelines: a comparison of two methods	No priority indicators were found
Guidelines and Protocols Advisory Committee	2017	GPAC Handbook	No priority indicators were found
Haller	2015	A survey on the methodological processes and policies of renal guideline groups as a first step to harmonize renal guidelines	No priority indicators were found
Handoll	2013	A framework for effective collaboration between specialist and broad-spectrum groups for delivering priority Cochrane reviews	Seems to cover priority-setting of new topics
Haute Autorité de Santé	2016	Élaboration de recommandations de bonne pratique.	No priority indicators were found
Hoomans	2012	Systematizing the Use of Value of Information Analysis in Prioritizing Systematic Reviews [Internet]	Seems to cover priority-setting of new topics
Italian Society for Haemostasis and Thrombosis	2009	Objectives and methodology: Guidelines of the Italian Society for Haemostasis and Thrombosis (SISET).	No priority indicators were found
Javaher, S. P.	2015	Guideline Development Process in a Public Workers' Compensation System	No priority indicators were found
Joanna Briggs Institute Synthesis Science Unit		Best Practice Information Sheet (BPIS) Procedures.	Could not be found online
Jones	2013	Success for a novel approach to priority setting in South Australian public dental clinics	Health care priority setting
Kim	2018	Identifying and prioritizing topics for evidence-based geriatric nursing practice guidelines in Korea	Does not seem to cover prioritization for updating

Kwaliteitsinstituut voor de Gezondheidszorg CBO	2007	Evidence-based Richtlijnontwikkeling Handleiding voor werkgroepleden.	No priority indicators were found
Lyratzopoulos	2012	Updating clinical practice recommendations: is it worthwhile and when?	No priority indicators were found
Martinez Garcia	2012	Strategies for monitoring and updating clinical practice guidelines: a systematic review	No original data
Martinez Garcia	2014	Updated recommendations: an assessment of NICE clinical guidelines	No priority indicators were found
Martinez Garcia	2017a	Continuous surveillance of a pregnancy clinical guideline: an early experience	No priority indicators were found
Martinez Garcia	2017b	Methodological systematic review identifies major limitations in prioritization processes for updating	No original data
McClarey	1999	Identifying priorities for national clinical guidelines	Seems to cover priority-setting of new topics
Mickenautsch	2013	The modified Ottawa method to establish the update need of a systematic review: glass-ionomer versus resin sealants for caries prevention	No priority indicators were found
Moher	2008	When and how to update systematic reviews	No original data, duplicate in PubMed
Moher	2007	A systematic review identified few methods and strategies describing when and how to update systematic reviews	No original data
Moher	2008	When and how to update systematic reviews	No original data
Nasser	2013a	An equity lens can ensure an equity-oriented approach to agenda setting and priority setting of Cochrane Reviews	No priority indicators were found
Nasser	2013b	Ensuring relevance for Cochrane reviews: evaluating processes and methods for prioritizing topics for Cochrane reviews	No priority indicators were found
Nast	2019	Prioritizing topics in guideline development: results of a two-phase online survey of dermatologist members of the EADV	Unclear whether it covered prioritization for updating
National Health and Medical Research Council	2009	A guide to the development, implementation and evaluation of clinical practice guidelines.	No priority indicators were found
New Zealand Guidelines Group		Handbook for the Preparation of Explicit Evidence-base Clinical Practice Guidelines.	Could not be found online
NICE	2013a	Interim Process and Methods Guide for the Clinical Guideline Updates Using Standing Committees Pilot Programme 2013 [Internet]	No priority indicators were found
NICE	2013b	Interim Process and Methods of the Highly Specialised Technologies Programme [Internet]	No priority indicators were found

NICE	2014	Interim Methods Guide for Developing Service Guidance 2014 [Internet]	No priority indicators were found
NICE	2013c	Interim Clinical Guideline Surveillance Process and Methods Guide 2013 [Internet]	No priority indicators were found
Pattanittum	2012	A comparison of statistical methods for identifying out-of-date systematic reviews	No priority indicators were found
Peterson	2011	Decisions to update comparative drug effectiveness reviews vary based on type of new evidence	No priority indicators were found
Robinson	2015	Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies	No original data
Royal Dutch Society for Physical Therapy	2016	Guideline methodology manual 2016	No priority indicators were found
Sampson,	2008	Surveillance search techniques identified the need to update systematic reviews	No priority indicators were found
Scott	2018	Cochrane acute respiratory infections group's stakeholder engagement project identified systematic review priority areas	Seems to cover priority-setting of new topics
Scottish Intercollegiate Guidelines Network	2015	SIGN 50: A guideline developer's handbook.	No priority indicators were found
Shekelle	2011	Identifying Signals for Updating Systematic Reviews: A Comparison of Two Methods [Internet]	No priority indicators were found
Shekelle	2009	Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005–2009) [Internet]	No priority indicators were found
Shekelle	2001	Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated?	No priority indicators were found
Shojania	2007	Updating Systematic Reviews	No priority indicators were found
Shojania	2007	How quickly do systematic reviews go out of date? A survival analysis	No priority indicators were found
Society for Vascular Surgery	2011	Guideline methodology of the Society for Vascular Surgery including the experience with the GRADE framework.	No priority indicators were found
Soll	2008	Updating reviews: the experience of the Cochrane Neonatal Review Group	Seems to be a need for updating strategy
Taylor	2019	Use of multi-attribute decision-making to inform prioritization of Cochrane review topics relevant to rehabilitation	No priority indicators were found
Therapeutic Guidelines Limited	2017	How Therapeutic Guidelines are produced.	No priority indicators were found
Tsertsvadze	2011	Updating comparative effectiveness reviews: current efforts in AHRQ's Effective Health Care Program	No priority indicators were found

Tugwell	2013	Methods for setting priorities in systematic reviews	No priority indicators were found
Vandvik	2014	[A new generation of reliable clinical practice guidelines through MAGIC]	No priority indicators were found
Vernooij	2014	Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks	No original data
Voisin	2008	Strategies in assessing the need for updating evidence-based guidelines for six clinical topics: an exploration of two search methodologies	No priority indicators were found
Wale	2013	The Cochrane Library review titles that are important to users of health care, a Cochrane Consumer Network project	Does not seem to cover prioritization for updating
Welsh	2015	Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach	Seems to be a need for updating strategy
Working Group on CPG Updates	2009	Updating Clinical Practice Guidelines in the Spanish National Health System: Methodology Handbook.	No priority indicators were found

Table A2 – General characteristics of included studies

Author or organization†	Year	Country‡	Brief description of the priority-setting assessment	Outcome level	Ordinal categories or cut-off value
Agbassi et al.[1]	2014	Canada	Documents that were selected to be reviewed for currency and relevance were prioritized using indicators presented in a flow diagram. Priority reviews were then further assessed for currency.	Ordinal	Urgent / High / Medium / Low
Ahmadzai et al.[2]	2013	Canada	Conclusions from each specific key-question was assessed for prevalent outdatedness. A global priority score for updating was then assigned based on the proportion and extent of the outdatedness.	Ordinal	High / Medium / Low
Akl et al. [3]	2017	Lebanon	Three criteria indicate whether recommendations should be prioritized and assigned a living status.	NF¶	NF¶
American Academy of Orthopaedic Surgeons[4]	2011	United states of America	A committee will prioritize topics by answering seven questions about each guideline. Thereafter, voting will decide which guideline is pursued.	NF¶	NF¶

Jiang et al.[5]	2019	China	NF¶	NF¶	NF¶
National Institute for Health and Care Excellence[6]	2014	United Kingdom	The organization prioritizes according to the users' need for new and updated guidelines.	NF¶	NF¶
Newberry et al.[7]	2013	United States of America	An overall priority for updating judgement was made based on two indicators.	Ordinal	High / Medium / Low
Sanabria et al.[8]	2020	Spain	Key-questions are scored individually on seven priority indicators with a tool. Priority scores are calculated and ranked. A final priority decision is based on the item and priority scores among other indicators.	Ranked continuous	Cut-off: NF¶
Shekelle et al.[9]	2014	United states of America	An overall priority for updating judgement was made based on two indicators.	Ordinal	High / Medium / Low
Shekelle et al.[10]	2014	United States of America	An overall priority for updating judgement was made based on two indicators.	Ordinal	High / Medium / Low
Sutton et al.[11]	2009	United Kingdom	Several methods were presented where the outcomes could be ranked (e.g. in decreasing order).	Ranked continuous	Cut-off: NF¶
Synnot et al.[12]	2020	Australia	Participants in the priority-setting voted for their top-5 priority topics. Topics were mapped against the current portfolio of reviews and editorial indicators were used to make list of reviews to be developed or updated. Discussion led to the selection of five reviews.	NF¶	Cut-off: 5 reviews
Takwoingi et al.[13]	2013	United Kingdom	The estimated probability of conclusions changing after new studies to a meta-analysis could be ranked in descending order.	Ranked continuous	Cut-off: NF¶
US Preventive Task Force[14]	2015	United stated of America	Members and partner organizations could prioritize active topics for review in the next 12 to 18 months based on four indicators. A working group assigned a priority recommendation to the topics, which became final after voting.	Ordinal	High/ Moderate / Low
World Health organization[15]	2014	Switzerland	NF¶	NF¶	NF¶

†The first author or the organization was extracted.
‡The country mentioned in the first affiliation of the first author, or the country of the organization was extracted.
¶NF: Not found or could not be deduced.

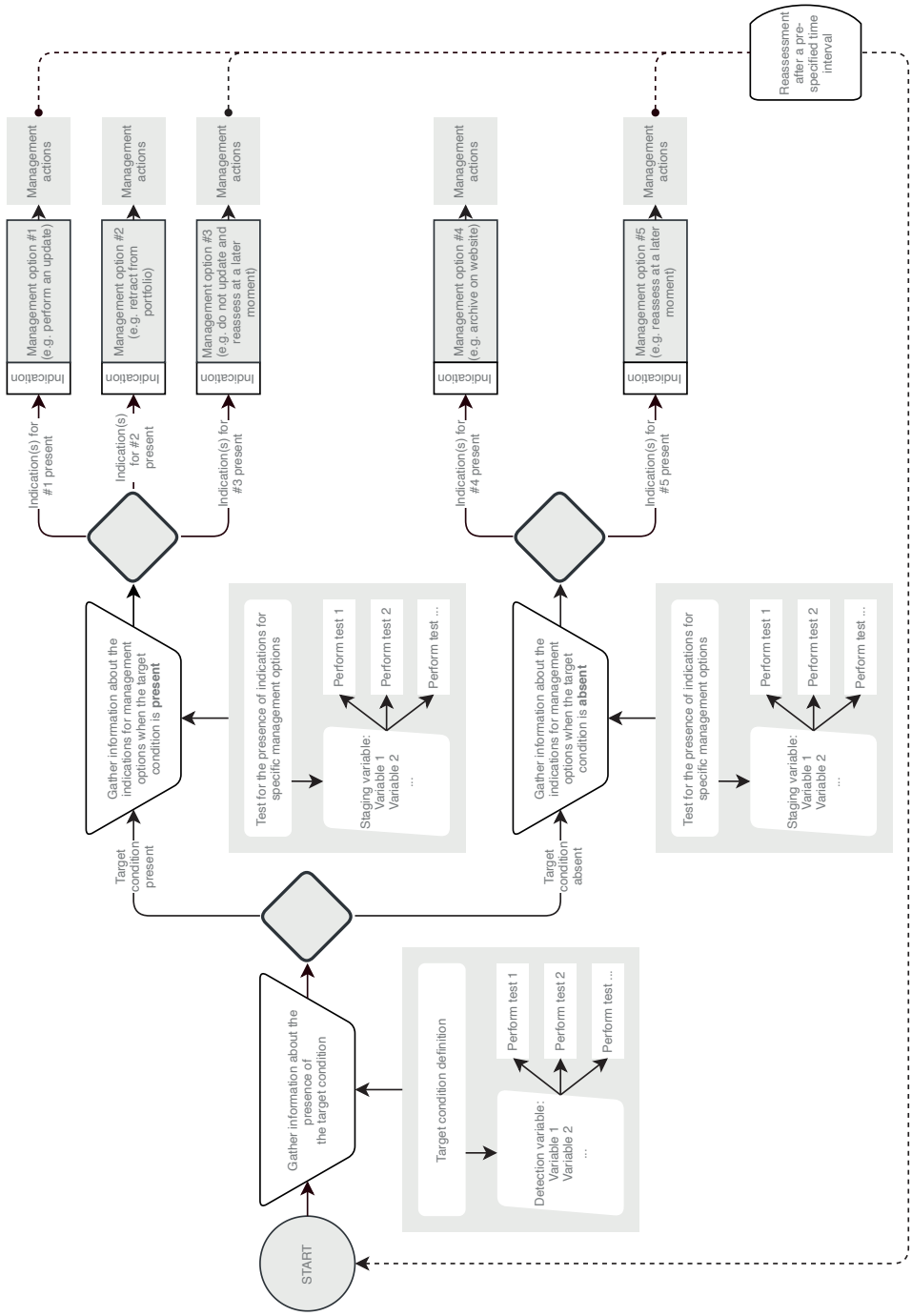


Figure A2 – The Portfolio Maintenance by Test-Treatment (POMBYTT) framework based on a diagnostic test-treatment pathway. *The depicted framework does not show priority-setting assessments.*

Table A3 – Priority-setting assessment description table

Priority-setting assessment	*Role*	<triage or add-on>	
	(Describe position in the strategy)	<position in the strategy, before/after which tests or actions>	
	Outcome level	<ordinal or ranked continuous>	
	(Describe levels if ordinal)	<ordinal levels if applicable>	
	(Describe cut-off / limit of capacity)	<cut-off determining what is high enough or not enough priority to continue>	
	Downstream consequence(s)	**High enough priority**	**Not enough priority**
		<downstream consequence of having a high enough priority>	<downstream consequence of having not enough priority>

Table A4 – A hypothetical example of a 'living' recommendations strategy with added priority-setting assessments, based on information found in Akl et al. 2017,[3] El Mikati et al. 2022,[16] and Bragge et al. 2022[17] which we adapted for illustrative purposes.

Priority-setting assessment	*Role*	Triage	
	(Describe position in the strategy)	Guideline recommendations are prioritized before they may enter the living updating strategy and continue for further testing.	
	Outcome level	Ordinal	
	(Describe levels if ordinal)	(Low priority / High priority)	
	(Describe cut-off / limit of capacity)	(Only high priority may continue)	
	Downstream consequence(s)	**High enough priority**	**Not enough priority**
		High priority recommendations continue to receive continuous monitoring	Guideline recommendations with not enough priority are send back to the portfolio for a re-assessment of their priority at a later point in time

Diagnosis	*Target condition definition*	The guideline recommendation is out of date when any new evidence becomes available
	Detection variable	New scientific evidence according to the existing literature selection criteria
	Detection test (protocol)	Literature search and selection (MEDLINE and EMBASE are searched by rerunning the previous search strategy. The search strategy was developed by an information specialist. The title and abstract of the retrieved hits are screened by two methodologists independently. Conflicts are resolved. The resulting full text articles are screened by two methodologists for inclusion. Data is extracted from the Included articles and study quality is appraised.)
	Detection test threshold	Any new relevant published evidence

		Target condition present		**Target condition absent**
Staging	*Staging variable*	Impact of new relevant published evidence on the recommendations	Impact of new relevant published evidence on the certainty of evidence	–
	Staging test (protocol)	Expert panel discussion and voting (The panel is supplied with the review question and guideline recommendation, with the new evidence and study quality, and with a set of questions to determine the influence of the new evidence on the recommendation. The panel discusses the possible implications in a physical meeting and votes whether the recommendation is impacted by the new evidence)	Certainty of evidence evaluation (The GRADE Summary of Findings table is updated by methodologists with information from the new evidence and performing new GRADE assessments)	–
	Staging thresholds	**Impacted:** when >75% of the panel indicates that the new evidence has impact on the recommendation **Uncertain impact:** 25-75% of the panel indicates that the new evidence has impact on the recommendation **No impact:** >25% of the panel indicates that the new evidence has impact on the recommendation	**High certainty:** when all critical outcomes for decision-making have a HIGH GRADE. **Relatively certain:** when most, but not all, critical outcomes for decision-making have HIGH GRADE **Moderate certainty:** when most or all critical outcomes for decision-making have MODERATE GRADE. **Uncertain:** When most critical outcomes for decision-making have LOW or VERY LOW GRADE	–

		Target condition present	Target condition absent
Management	*Management options (indications) and actions*	**Update the recommendation and continue monitoring** (when the expert panel indicates the recommendation is impacted by new evidence [>75%]) Perform an update using the regular procedures	**Do not update and continue monitoring** (When no new relevant evidence was found) Continue monitoring as defined under 'Monitoring'
		Do not update and continue monitoring (When the expert panel was uncertain [25-75%] OR indicated that there was no impact [<25%]) Continue monitoring as defined under 'Monitoring'	–
		Update the recommendation and retire from living status (when the expert panel indicates the recommendation is impacted by new evidence [>75%] AND the certainty of evidence became high) Perform an update using the regular procedures and retire the recommendation from participating in this maintenance strategy.	–

Priority-setting assessment	*Role* *(Describe position in the strategy)*	Add-on Guideline recommendations are prioritized after it is decided they need an update when the certainty did not become HIGH, but before they actually receive the update when there are more updates to perform than there is capacity	
	Outcome level *(Describe levels if ordinal)* *(Describe cut-off / limit of capacity)*	Ranked continuous (–) (There are resources to perform 5 updates, thus the limit of capacity is set at five recommendations)	
	Downstream consequence(s)	**High enough priority**	**Not enough priority**
		The five top-ranked recommendations receive an update	Guideline recommendations with not enough priority are re-assessed whether the living status still applies.

Monitoring	*Monitoring*	Weekly reruns of the existing search strategy

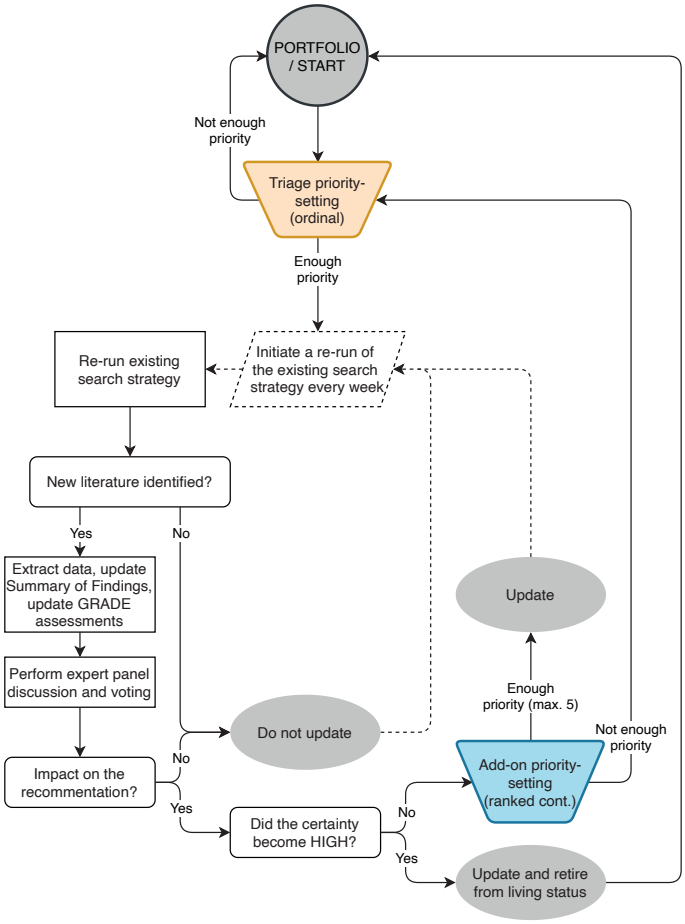


Figure A3 – Process flow diagram of the hypothetical 'living' example strategy with added priority-setting assessments *(see Table A4) based on information found in Akl et al. 2017,[3] El Mikati et al. 2022,[16] and Bragge et al. 2022[17] which we adapted for illustrative purposes. We have added a second priority-setting assessment (blue) at our own discretion for illustrative purposes.*

LITERATURE LIST

1. Agbassi C, Messersmith H, McNair S, Brouwers M. Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. J Clin Epidemiol. 2014;67(12):1335-42.
2. Ahmadzai N, Newberry SJ, Maglione MA, Tsertsvadze A, Ansari MT, Hempel S, et al. A surveillance system to assess the need for updating systematic reviews. Syst Rev. 2013;2:104.
3. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol. 2017;91:47-53.
4. American Academy of Orthopaedic Surgeons Clinical Practice Guideline Unit. Guideline Process Procedure for Updating Existing Evidence-Based Clinical Practice Guidelines. American Academy of Orthopaedic Surgeons; 2011.
5. Jiang ZM, Zhan SY, Zuo L, Jia XW, Fang H, Li XX, et al. Guideline for [Clinical Guidelines Constitution/Amendment] in China. Int J Qual Health Care. 2019;31(7):568-74.
6. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual (April 2017 update). 2014.
7. Newberry SJ, Ahmadzai N, Motala A, Tsertsvadze A, Maglione M, Ansari MT, et al. AHRQ Methods for Effective Health Care. Surveillance and Identification of Signals for Updating Systematic Reviews: Implementation and Early Experience. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
8. Sanabria AJ, Pardo-Hernandez H, Ballesteros M, Canelo-Aybar C, McFarlane E, Niño de Guzman E, et al. The UpPriority tool was developed to guide the prioritization of clinical guideline questions for updating. J Clin Epidemiol. 2020;126:80-92.
9. Shekelle PG, Motala A, Johnsen B, Newberry SJ. Assessment of a method to detect signals for updating systematic reviews. Syst Rev. 2014;3:13.
10. Shekelle PG, Motala A, Johnsen B. AHRQ Methods for Effective Health Care. Assessment of a Method to Detect Signals for Updating Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
11. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. J Clin Epidemiol. 2009;62(3):241-51.
12. Synnot AJ, Tong A, Bragge P, Lowe D, Nunn JS, O'Sullivan M, et al. Selecting, refining and identifying priority Cochrane Reviews in health communication and participation in partnership with consumers and other stakeholders. Health Res Policy Syst. 2019;17(1):45.
13. Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. Bmj. 2013;347:f7191.
14. US Preventive Services Task Force. US Preventive Services Task Force Procedure Manual. 2015.
15. World Health Organization. WHO handbook for guideline development - 2nd edition. Geneva, Switzerland: WHO Press; 2014.
16. El Mikati IK, Khabsa J, Harb T, Khamis M, Agarwal A, Pardo-Hernandez H, et al. A Framework for the Development of Living Practice Guidelines in Health Care. Ann Intern Med. 2022;175(8):1154-60.

17. Bragge P, Allison DJ, Mehta S, Guy S, Loh E. A pilot living clinical practice guidelines approach was feasible and acceptable to guideline panel members. J Clin Epidemiol. 2022;143:22-9.

CHAPTER 6

Introducing re-weighted range voting in clinical practice guideline prioritization: development and testing of the RE-weighted Priority-Setting (REPS) tool

M.S. Oerbekke
C.M.W. Gaasterland
M.J. van der Laan
L. Hooft

PLoS One. 2024 Apr 5;19(4):e0300619
doi: 10.1371/journal.pone.0300619
PMID: 38578723

ABSTRACT

We aimed to develop and test a tool based on the re-weighted range voting mechanism to prioritize items (i.e. key questions) in a priority-setting assessment for clinical practice guidelines. The secondary aim was to provide methodological context of the tool. We iteratively developed the tool and used qualitative methods (i.e. think-aloud and semi-structured interviews) to test the tool's usability and make adjustments accordingly. An observational approach was used to test the tool's outcome satisfaction in a real-world priority-setting assessment within a rare-disease guideline of a European Reference Network and under four different conditions in the tool. Four guideline methodologists tested the usability of the tool. The real-world testing was performed with a guideline panel consisting of a core working group, five expertise working groups, and a working group with patient representatives. Thirty-one panel members assigned scores in the priority-setting assessment. Seventeen panel members rated the priority-setting outcome, and sixteen panel members rated the outputs generated under the four conditions. Upon initial use, guideline methodologists found the tool to be quite overwhelming. However, with some initial effort they were able to easily identify the tool's structure. Based on observations and feedback, the tool was further refined and user guidance was developed. Guideline panel members expressed (high) satisfaction with the priority-setting outcome. They particularly preferred the condition when using mean subgroup scores as input or employing aggressive penalties in the weighting method to determine the outputs. The tool generates a ranked list of items and offers flexibility for different choices in priority-setting assessments as long as its input format requirements are met. Although it is not a consensus method, the tool assists in narrowing down a set of priority items. Additional steps in the priority-setting assessment can lead to a consensus being reached regarding the final outcome.

1. INTRODUCTION

Clinical practice guideline (CPG) developing organizations often cope with limited resources, which can result in a mismatch between the required work (e.g. updating) and the available capacity. Such limited resources may be a driver for the need to identify key items that require the allocation of resources above other items. Priority-setting assessments aim to help identifying such items (e.g. key questions, recommendations). By prioritizing these items, organizations can allocate their limited resources more effectively. Priority-setting is therefore a crucial process for CPG developing organizations who seek to optimize their resource utilization.

Multiple systematic reviews were conducted synthesizing priority-setting processes and indicators,[1-3] including those related to the priority-setting of CPG topics,[4] key questions in CPG development,[5-7] CPG recommendations for uptake,[8] and implementation.[9] Although it was stated that there is a need to develop standardized and validated priority-setting tools[5] and some methods have both been developed[6, 10, 11] and tested,[7, 12, 13] it is important to recognize that different CPG developing organizations may have unique needs and operate in different contexts. No single published priority-setting assessment might therefore be fully generalizable to other CPG developing organizations without adaptation. Instead, it may be beneficial to develop methodological knowledge for priority-setting assessments and offer multiple options within priority-setting assessments. From here, organizations may choose steps, processes and indicators,[1-3] and tools or methods[6, 7, 10-13] that best aligns with their purpose and context. That is, organizations may choose which steps (e.g. generation of topic list, collect data to inform discussions, use of prioritization criteria, ranking, refinement[1]) their assessment should contain, how these steps should be conducted (e.g. using stakeholder input or use existing CPGs to generate a topic list[1]), and which priority indicators (e.g. health burden[1, 2] or burden of disease,[3] practice variation[3]) are deemed relevant. Such an approach could facilitate more effective utilization of the limited resources for CPG development within organizations due to the priority assessments being self-tailored to their own context, including their available resources to conduct the priority-setting assessment. Tools and methods suggesting a set of priority indicators may be at risk of offering indicators irrelevant to the context of organizations. For example, using 'new available evidence' and 'consequences for costs' as priority indicators may not necessarily seem to apply in dental care.[7] Furthermore, imposed steps and processes may cost resources (in a broad sense) that organizations might not be willing to invest or ask from participants. For example, for 107 questions rated by 30 participants on multiple indicators it took an average of 3.8 hours per participant[13] and thus an average total of 114 hours. In perspective, this is almost 10 percent of a 40-hour workweek for an individual participant and a total almost equaling the weekly productivity of three full-time employees. Some organizations might feel such a process is too time-intensive or puts too much strain on the participants and may wish to use a less resource-intensive process (e.g. less priority indicators to score, providing only one overall score, or even use a method without priority indicators).

Priority-setting assessments have relied on Delphi methods,[4, 5, 8] counting,[14] or calculating

mean scores.[6, 11, 13] However, some contexts may need other options for initially ranking or selecting items. For example, results may be distorted when discussions involve persuasive actors or the composition of the participants is imbalanced. In a multidisciplinary oncology CPG, for example, there may be more representatives from internal medicine and surgery in the CPG panel than compared to pathology. This could lead to outcomes favoring the interests of the largest groups. However, pathology is critical for diagnosing and staging the disease to initiate a systemic and/or surgical treatment. New developments in this area may be important but not prioritized through counting or averaging due to imbalances in the representation. This could leave clinical guidance sub-optimal. Therefore, there is a need to explore additional options which could help less represented perspectives being considered in priority-setting assessments.

Re-weighted range voting uses a mechanism that adjusts the influence of participants for selecting the next item based on their scores on previously selected items.[15] This mechanism has a proportional representation characteristic,[15] but can be modified to enhance the representation of less dominant perspectives in the ranking. This could depend on which data is used and the specific modifications made to the mechanism. However, before such tool can be actually used, it first needs to be developed and tested in real-world scenarios. Our goals were to develop a tool based on the re-weighted range voting mechanism (including a modified mechanism), to evaluate its usability, and to evaluate its outcome satisfaction in a real-world priority-setting assessment. Additionally, we aimed to provide methodological context for the tool to help CPG developing organizations determine whether our tool is appropriate for their priority-setting purpose and contextual requirements.

2. MATERIALS AND METHODS

2.1 ETHICS STATEMENT

Review by a Medical Research Ethics Committee in the Netherlands is not required when the study is not under the scope of the Dutch Medical Research involving Human Subjects Act,[16] where the current type of study does not meet the two conditions needed to fall under its scope.[17]

2.2 FRAME OF REFERENCE

A theoretical frame of reference was developed to discern key components of priority-setting assessments. Through a creative process, ideas pertaining to priority-setting assessments were integrated and adapted to generate abstract main components. The frame of reference will show the concept of sequential steps in a priority-setting assessment based on the generated main components and aimed to illustrate how we considered the role of our RE-weighted Priority-Setting (REPS) tool in priority-setting assessments to be. Such steps and their procedures may be dependent of the context and wishes of an organization. For example, an organization may choose to generate a topic list as an early step in their

priority-setting assessment by surveying experts. To establish the initial face validity of the frame of reference, several priority-setting assessments published in the literature were aligned with its components.

2.3 DEVELOPMENT OF THE REPS-TOOL

2.3.1 RE-WEIGHTED RANGE VOTING MECHANISM

Participants are instructed to assign scores from 0 (i.e. no priority) to the maximum scale score (i.e. highest possible priority; determined a priori). Participants also have the option to abstain from scoring an item, while negative scores are not allowed. The item with the highest total score is selected as the first winner.[15] The individual weight of participants are then adjusted based on their scores assigned to previously selected winners, using Formula 1:[15, 18]

$$(1)\quad individual\ weight = \frac{constant}{(constant + \frac{sum\ of\ participant's\ scores\ on\ ranked\ winners}{maximum\ scale\ score})}$$

The participant's scores on the remaining items are multiplied by their respective individual weights and the new scores are summed for each item. The highest total score is subsequently identified as the next winner. We additionally used a second weighting method to achieve a more disproportionate representation by introducing a parameter (here denoted as *A*) as an exponent of the participant's sum of scores on ranked winners (Formula 2).[15] We will refer to this method as the decay-adjusted weighting method, and to the exponent as decay aggression (*A*):

$$(2)\quad individual\ weight = \frac{constant}{(constant + \frac{sum\ of\ participant's\ scores\ on\ ranked\ winners^{A}}{maximum\ scale\ score})}$$

2.3.2 EARLY USABILITY TESTING

Think aloud sessions[19] with semi-structured interviews were performed with CPG methodologists, who were invited in June-July 2020 and agreed to participate. The methodologists participated in individual video calls where they utilized a preliminary version of the REPS-tool. The methodologists were informed about the purpose, the session, the recording (including the anonymized transcription), and were provided with the opportunity to ask any questions (see Additional File 1 for the protocol). The methodologists were asked to share their screen and communicate their thoughts out loud while they were performing tasks in the REPS-tool during the session. One author (MSO) led the one-to-one sessions and recorded the video call for later analysis. The recorded videos were transcribed word-by-word in Dutch and enriched with descriptions of the methodologists' on-screen actions. One author (MSO) analyzed the transcriptions by hand and added qualitative labels to observed experiences in Dutch (labels were also translated to English whilst quotes remained in Dutch). The observations were thereafter organized in clusters based on the different phases in the think aloud sessions, forming the basis for subsequent interpretations.

2.3.3 REAL-WORLD PRIORITY-SETTING

We used the REPS-tool in the priority-setting of key questions in the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability (ERN-ITHACA) rare-disease CPG concerning the Kleefstra syndrome (Fig 1). There were resources to newly develop 12 out of 45 key questions defined by the CPG panel. The panel consisted of a core group (including a CPG methodologist), five working groups (based on expertise), and a working group of patient representatives. Three priority indicators were selected by the core working group: 1) unwanted (international) clinical variance that may be improved with the guideline, 2) high prevalence of symptoms, and 3) high (disease) burden for either the Kleefstra population or for their relatives. Panel members were invited to an online survey and were asked to individually provide an overall priority score ranging from 0 to 5 per key question based on the three indicators. The mean scores for each key question within each working group were entered into the REPS-tool. The regular re-weighted range voting (Formula 1) was used as the weighting method. Thereafter, the resulting top-18 rankings were discussed by the CPG panel's core group. This led to the selection of 12 key questions as the final outcome of the priority-setting assessment.

The set of priority scores from the survey was also used to generate top-10 outputs of the REPS-tool, considering four different conditions in order to examine their impact on the outcome: 1) mean scores from the working group while using the regular re-weighted range voting, 2) individual participant scores while using the regular re-weighted range voting, 3) individual participant scores while using a low decay aggression, and 4) individual participant scores while using a high decay aggression. During an online meeting, the CPG panel members (excluding the core group) individually indicated their preferred top-10 outputs through a poll. CPG panel members were thereafter presented with the core group's selection of 12 key questions and were asked to express their satisfaction level with this final outcome of the priority-setting assessment (ordinal poll with five response levels: very content, content, medium, not very content, and dissatisfied).

3. RESULTS

3.1 FRAME OF REFERENCE

Three main components in priority-setting assessments were discerned (Fig 2): the process component, the function component (including its input and output), and the outcome component. The process component involves procedural steps followed in the priority-setting assessment. The function component contains specific rules and/or calculations that determine the priority of items. It requires an input to perform its function, and results in an output. The outcome component represents the final outcome of the priority-setting assessment. It can directly correspond to the output of a function (Fig 2A), or additional process components are added to achieve a final outcome of the priority-setting assessment (Figs 2B and 2C). The initial face validity of the frame of reference was examined by aligning priority-setting assessments from the literature to its components (Additional File 2).

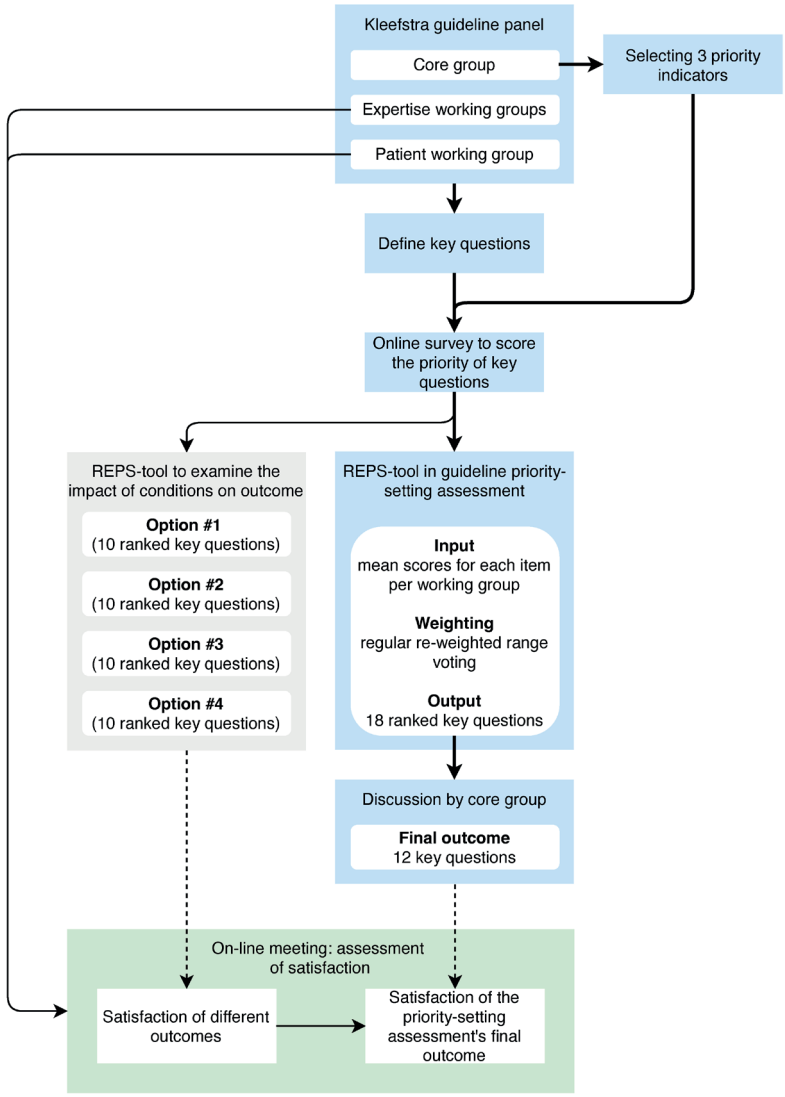


Fig 1. Flow of the priority-setting assessment pilot. *The figure shows the process of the priority-setting assessment in the Kleefstra guideline (blue boxes and bold arrows). The assigned scores in the survey were also used to test four different conditions in the REPS-tool determining the outcome (grey box). Satisfaction of the four different outcomes of the REPS-tool (from the grey box) and the final outcome of the Kleestra priority-setting assessment (from the blue process) were assessed in an online meeting (green box).*

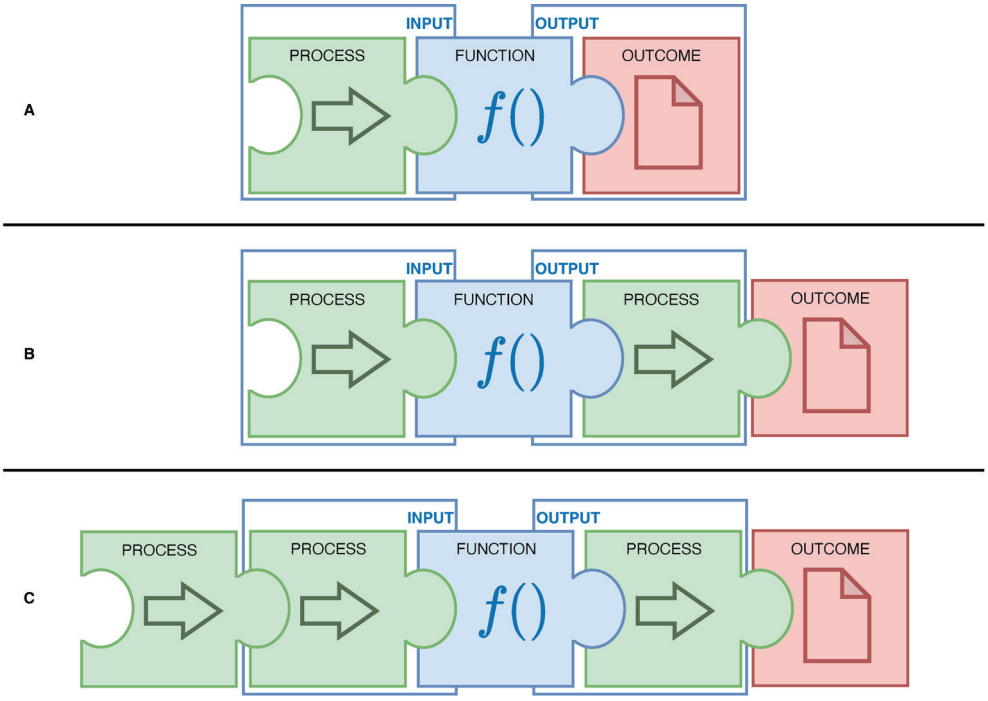


Fig 2. Priority-setting assessment components. *The figure shows three main components of a priority-setting assessment: process, function, and outcome. In a basic form, a single process component leads to input for the function which produces an output as the final outcome of the priority-setting assessments (Fig 2A). Priority-setting assessments can be extended by incorporating additional process components tailored to the specific needs of an organization. For example, adding a process component to discuss the function's output to determine the final priority-setting outcome (Fig 2B and 2C). Alternatively, multiple process components can be introduced before the function component (e.g. to select participants and to identify relevant priority-indicators) leading to the function's input (Fig 2C).*

3.2 EARLY USABILITY TESTING

All four participating CPG methodologists who participated were female, with their ages ranging from 26 to 32 years old. Their overall working experience varied from 3 to 8 years, with 1 to 3 years specifically in CPG development. Among the methodologists, three held a PhD degree and one an MSc degree. Three of the methodologists had experience with priority-setting in some way, of which one of them had seen the preliminary tool previously before participating in the usability test.

The first impression of the REPS-tool was that it was large and complex. The tool's structure concerning the placement of scores and ranks was initially unclear. Participants became familiar with the structure after some experimenting, through performing the assignments, or some verbal instructions. Once the structure was clear, we observed that the participants placed the input data correctly in the tool and could get the output out of the tool. Participants generally thought the automated functions in the tool were convenient and wished for more

automation in other parts of the tool, for example, one participant was missing a feature that would check for valid scores. The manual ranking process was more inconvenient when there were a lot of items in the tool. This was mainly due the difficulty of searching for the identified winner in the worksheet and remembering which rank to assign to the winner. Furthermore, the participants preferred to have some background information on the tool's mechanism and receiving guidance about how to use the decay aggression effectively. The qualitative analysis and all actions to improve based on this usability testing can be found in Additional File 3.

3.3 THE REPS-TOOL FUNCTION COMPONENT

The REPS-tool can be found in Additional File 4, programmed as an Excel-file. It was designed as a function component in a priority-setting assessment (Fig 3). Its input format is a matrix containing priority scores per participant per item, where rows represent participants (or subgroups) and columns represent items. The tool can highlight input outside the allowed range to validate the input and is capable of processing 100 items for 100 participants. A (subsequent) winner is automatically identified to which a rank can be semi-automatically assigned. To tool also identifies when and how many items are tied for a rank. Heterogeneity analyses within the REPS-tool could be decisive for which of the items to assign the tied rank to. Either the regular re-weighted range voting (Formula 1) or the decay-adjusted method (Formula 2) can be selected in the tool. In the latter, the decay aggression (*A*) is controlled in the tool's parameter section. The decay adjusted method places additional emphasis on the participant's first few assigned scores to the ranked winners. Fig 4 shows the pattern of the decay-adjusted individual weight depending on the value of *A*. Items with an assigned rank are ordered in a list by their rank and this output can be collected from the REPS-tool. More background information and guidance about how to use the REPS-tool can be found in the quick-start guide (Additional File 5).

3.4 REAL-WORD PRIORITY-SETTING

The priority-setting assessment in the Kleefstra syndrome CPG is displayed in an online supplementary file (A6 Fig 1 in Additional File 6), together with the assigned scores in an anonymized file (Additional File 7). Thirty-one panel members (including 10 patient representatives) assigned priority scores in the survey. The priority-setting assessment led to the final selection of 12 key questions by discussing the REPS-tool's top-18 output (A6 Fig 2 in Additional File 6). Discussion resulted in merging four questions with other questions in the top-18 due to being similar topics. Three questions, ranked 13th, 15th, and 17th in the top-18 output, were not selected. One question, initially not ranked in the top-18, was added through discussion. All 17 panel members, thus excluding the core group, voted in the poll regarding the priority-setting outcome. Eight members were 'very content' and nine were 'content' with the final priority-setting outcome as selected by the core group.

Four top-10 outputs were generated under different conditions (A6 Fig 3 in Additional File 6). Six of the same key questions remined in the tool's output regardless of the condition, albeit different ranks were assigned to these questions in each respective condition. Seventeen

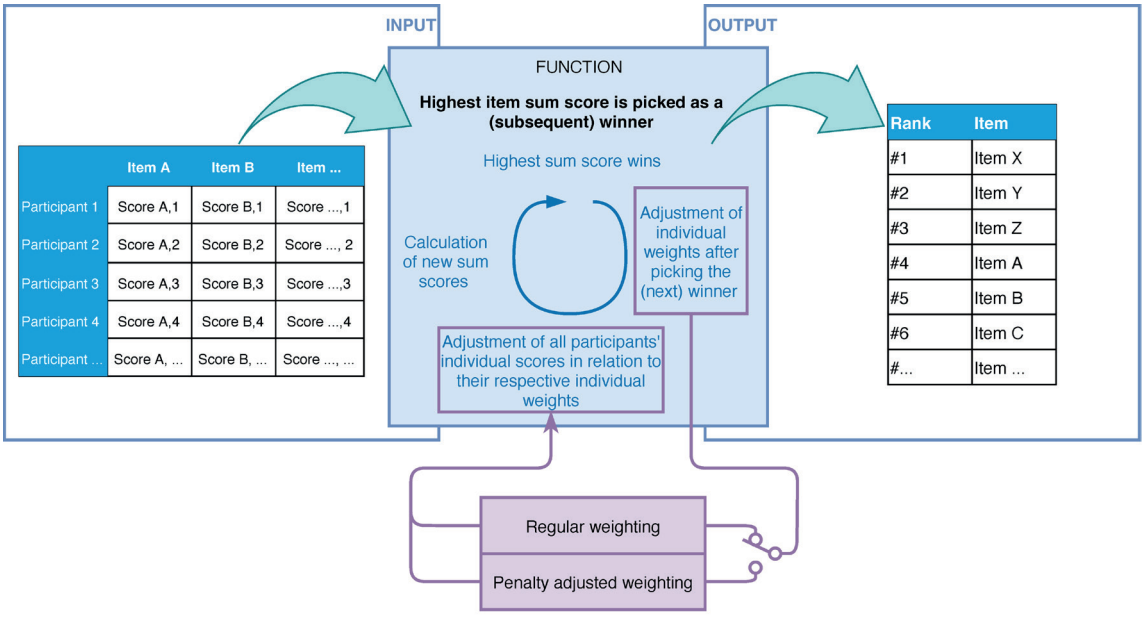


Fig 3. The REPS-tool as a function component. *The figure shows the REPS-tool as a function component. The input section shows the input format for the tool. The function repetitively identifies the (new) highest total score as a winner and assigns subsequent ranks to winning items. Individual weights are adjusted after assigning a rank to a winning item. Scores are thereafter recalculated based on the individual weight. The tool has two weighting options: regular re-weighted range voting and decay-adjusted re-weighted range voting. New item sum scores are calculated and the highest total score is identified as the next winner. The tool's output is a list of ranked items.*

6

panel members besides the core group participated in the online meeting where both the four top-10 outputs and the priority-setting outcome were evaluated. Sixteen panel members indicated which of the four top-10 outputs they preferred. Six panel members preferred the output generated from using scores of individual participants with a more aggressive decay (Option 4 from A6 Fig 3 in Additional File 6). Five panel members preferred the output obtained from the mean scores of sub-groups with regular re-weighted range voting (Option 1 from A6 Fig 3 in Additional File 6). Three panel members preferred the top-10 resulting from using the scores of individual participants with regular re-weighted range voting (Option 2 from A6 Fig 3 in Additional File 6). Using the scores of individual participants and a mild decay was preferred by 2 panel members (Option 3 from A6 Fig 3 in Additional File 6). The two most preferred options are compared in an online supplementary file (A6 Fig 4 in Additional File 6).

4. DISCUSSION

We developed a tool based on the re-weighted range voting mechanism which serves as a function component for CPG priority-setting assessments. While this function does not aim to achieve consensus, it can help narrowing down to a set of priority items by providing a ranked output. Additional process components within the priority-setting assessment can contribute

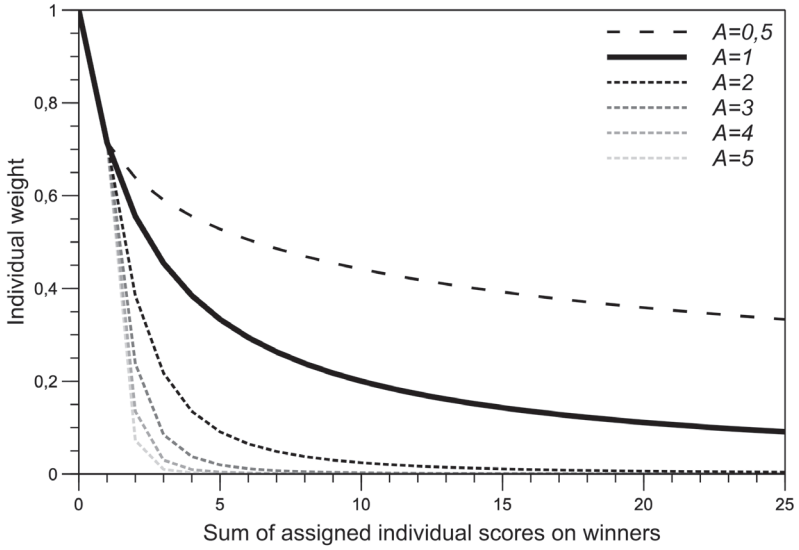


Fig 4. A graph showing the decay patterns of the individual weight. *The decay-adjusted weighting method controls the decay pattern of the individual weights. Any number provided for A (i.e. decay aggression) larger than 1 results in a more aggressive initial decay of the individual weight when compared to the regular re-weighted range voting method (A=1). Using A=0.5 will ease the decay of the individual weights in comparison to the regular re-weighted range voting method. Weights in the graph were calculated using constant of 0.5 and a maximum scale score of 10.*

to achieving consensus (e.g. discussing the tool's output with the panel to determine a final outcome). The tool underwent user testing during its development phase. Based on the findings from testing and interviews, adjustments were made to increase the usability and a quick-start guide was developed. Thereafter, a real-world priority-setting assessment was performed to prioritize CPG questions using the REPS-tool as a function component. GPG panel members were (very) content with the final selection of 12 key questions resulting from the priority-setting assessment. Parallel to the real-world priority-setting assessment, we had the opportunity to examine various input formats and weighting methods to assess four hypothetical top-10 outputs generated by the REPS-tool. Interestingly, among the sixteen voting panel members only three preferred the use of individual scores with regular re-weighted range voting, which provided proportional representation. The most favored options were using mean sub-group scores as input with the regular re-weighted range voting weighting method, or using the individual scores with an aggressive decay (i.e. *A*=4). These preferences warrant further examination and theoretical analysis of these conditions.

Using the mean scores of subgroups with regular re-weighted range voting, the proportional representation can be achieved at the subgroup level. This approach determines the ranked winners based on the priority assigned by subgroups, regardless of their size. This may be why panel members in the Kleefstra CPG identified this method as one of the two preferred options. However, it is not always feasible to predefine or differentiate subgroups. In such cases, obtaining mean scores per subgroup becomes impractical and individual

scores are used as input instead. As an alternative option, an aggressive decay can be applied to quickly decrease the individual weights of participants. Those participants who had assigned (high) scores to the first few winners have their influence quickly reduced for identifying the subsequent winners. This could leave some room for smaller subgroups or individuals with high individual weights to have their priority items receive a higher ranking. However, there is a risk that a large group of participants with moderate or low individual weights might still outweigh a small group with high individual weights. Thus, theoretically, a more aggressive decay is needed to significantly reduce the influence of the participants initially 'having their way' with the first few winners and accommodate for smaller subgroups and individuals. This might clarify why an aggressive decay adjusted weighting method was the second of the two preferred options in the Kleefstra CPG priority-setting. It is therefore important to think about how the group of participants is composed and whether potential subgroup formation or unbalanced representation is considered (un)desirable. Subgroup formation or unbalanced representation might be desirable in some situations. For example, when the priority-setting assessment takes place in a more mono-disciplinary setting. A larger number of persons from one or a few selected disciplines will probably participate, because they will be affected most by the recommendations. One can argue that they then would need to have the most influence in the priority-setting assessment too. Theoretically, the regular re-weighted range voting mechanism could be used when subgroup formation or unbalanced representation is not considered problematic because of the mechanism's proportional representation characteristic. However, subgroup formation or unbalanced representation might be undesirable in other situations. For example, when situations arise where the number of representatives within a subgroup might undesirably distort the outcome of the priority-setting assessment, leaving the important perspectives of small subgroups unexposed in the outcome. One could argue that smaller groups have important perspectives on priority too, but these are potentially underexposed due to a low number of representatives participating in the priority-setting assessment. For example, the patient representation during a priority-setting assessment for clinical practice guidelines can be smaller than the representation of clinicians, but they do have important perspectives on where their priorities lie. Such situations could also arise from practical limitations, as some disciplines may struggle to deliver representatives for all of the ongoing guideline developments and updates, let alone enough representatives to provide some counterweight in priority-setting assessments against large, well-represented disciplines. It seems that the decay-adjusted weighting method could be used to enhance less represented perspectives, although it remains challenging to determine which exact decay aggression should be used.

The REPS-tool uses a single priority score per participant per item. However, priority-setting assessments or tools described in literature seem to use multiple indicators.[6, 7, 11, 13, 14] When desirable, multiple scores could be averaged to comply to the input format requirement for the REPS-tool. A single overall score, however, might be sufficient for priority-setting in general,[6, 7] although scores on individual indicators may help to resolve ties.[6] This could also be helpful for resolving ties in the REPS-tool. Furthermore, some basic heterogeneity analyses are available in the REPS-tool that can also assist in making such decisions. Another helpful tool for CPG question priority-setting is the UpPriority Tool, which contains six priority indicators.[11, 13] It furthermore contains guidance for applying the tool, rating the indicators, calculating scores, ranking scores, the priority decision, and reporting the outcome.[11] The

REPS-tool does not contain any priority indicators or process components. Instead, users need to select their own process components (including the priority indicators if desirable) for their priority-setting assessment. This feature can be important in order to create or adapt assessments to the context of the priority-setting assessment when utilizing the REPS-tool.

Organizations performing priority-setting assessments may have different needs and operate within different contexts. Priority-setting assessments for CPGs in the context of venous thromboembolism[6] and COVID-19 dental care[7] showed differences in which indicators were significant predictors although using identical priority indicators. This could suggest that priority indicators might be dependent on the context,[7] for example, the healthcare system an organization operates in. Using 'the impact on the access to care' as a priority indicator[11, 13, 20] may be less discriminating for organizations operating in socialistic healthcare systems, which allow a high degree of access by default. Also, the aim of priority-setting can be different. Existing CPG recommendations or sections can be prioritized for uptake in other CPGs,[8] for implementation[9] or for updating.[12] All of such aims might require a different set of relevant priority indicators. Interestingly, literature shows that authors do select priority indicators based on the context, for example, of their national clinical practice.[14] Our frame of reference may aid in how priority-setting assessments are structured, while the steps, phases, and priority indicators described in literature[1] could guide decisions for process components regarding the organization's needs and context within this structure. If an organization would consider implementing the REPS-tool, it would need to recognize that the tool is solely a function component (see Figure 3) in a priority-setting assessment (see Figure 2) and not a priority-setting assessment in itself. Organizations may self-tailor their priority-setting assessment as long as the process components in the priority-setting assessment lead to input that complies to the REPS-tool's input format and when the organization desires a ranked list of items as output. Adaptations to the weighting methods can be made as all formulas and scripts are accessible and editable in the Excel-file when the current methods are not desired.

The theoretical nature of the REPS-tool as a function component is one of the limitations of the current study. Although we performed usability tests and piloted the tool in a real priority-setting assessment, there is still a need for a stronger evidence base regarding its usability. This includes determining which input format, weighting method, and decay aggression to use based on different contexts. We encourage others to use, test, compare, or adapt the REPS-tool and share their experiences. This collaborative effort will contribute to expanding the evidence base and help determine if the REPS-tool should be used in different contexts. We furthermore refined the tool based on the qualitative findings of the think aloud usability tests and semi-structured interviews. We did not perform a final check with the methodologists whether the adjustments addressed their usability matters and assumed the refinements were appropriate. Future plans involve a comprehensive evaluation of the tool's usability, applicability, and viability in real-world assessments. However, at face value, most matters may seem to be addressable by providing background information about the tool (which the methodologists deliberately did not receive for the usability testing at the time) and by further automating the tool.

Future development of the REPS-tool could include development of the tool in software

or apps linked to CPG databases for priority-setting in an online portal, including options to select different priority indicators and to elicit priority scores from stakeholders. Our pilot test focused on the satisfaction with the final outcome of a priority-setting assessment using the REPS-tool and four different outputs. Thorough evaluations of priority-setting assessment using the REPS-tool could be performed according to a pre-defined framework in the future [21]. Furthermore, an adapted version of the REPS-tool (not published) is currently being tested in-house at the Knowledge Institute of the Dutch Association of Medical Specialists with their stakeholders to assign priorities of CPG sections for updating as part of a continuous maintenance strategy. The use of the REPS-tool is also being considered in a second ERN-ITHACA CPG to select key questions for development after being well-received by the Kleefstra CPG panel. Furthermore, the tool could also be tested used outside the context of CPG development to prioritize any other item (e.g. knowledge gaps, patient reported outcome measures, systematic review questions, etc.). Hopefully, experiences with the REPS-tool and evaluations of priority-setting assessments using the tool will be shared.

5. CONCLUSIONS

The REPS-tool has the theoretical capability to rank various items, such as key questions, recommendations, and CPG topics to support priority-setting during the development or updating of CPGs. It could also have applications for priority-setting outside the field of CPGs, such as research questions, patient reported outcome measures, knowledge gaps, and more. As a function component in priority-setting assessments, it allows for a wide range of choices within the priority-setting assessment as long as its input format requirement is met (i.e. a single score per item per participant). For example, different sets of priority indicators can be selected based on the specific context of the priority-setting assessment. While the REPS-tool assists in the ranking of items during a priority-setting assessment, it should be noted that it is not a consensus method in itself. When using the REPS-tool, it is important to consider the composition of the participants in the priority-setting assessment and determine whether there are (potential) imbalances and whether they are desirable or not. These considerations may affect decisions regarding the score input (e.g. individual or mean subgroup scores) and the choice of weighting method (i.e. regular or decay-adjusted). However, more empirical evidence may be needed to support these choices. The REPS-tool helps narrowing down a set of items, while additional steps in the priority-setting assessment may be used to obtain consensus for a final outcome. It's worth noting that in the first real-world evaluation of priority-setting using the REPS-tool in CPG development, the panel members were (very) satisfied with the final outcome of the priority-setting assessment.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

All author and non-author contributions are made transparent using the structured Contributor Role Taxonomy (CRediT) described at https://doi.org/10.1002/leap.1210.

MSO: Conceptualization, methodology, software, investigation, formal analysis, data curation, writing – original draft, visualization
CMWG: Investigation, resources, writing – review & editing
MJvdL: Conceptualization, supervision, writing – review & editing
LH: Conceptualization, methodology, supervision, writing – review & editing

DATA AVAILABILITY STATEMENT

All relevant data are in the manuscript and its Supporting Information files. All formulas and VBA scripts in the REPS-tool's Excel file are visible, accessible, and editable.

LITERATURE LIST

1. El-Harakeh A, Lotfi T, Ahmad A, Morsi RZ, Fadlallah R, Bou-Karroum L, Akl EA. The implementation of prioritization exercises in the development and update of health practice guidelines: A scoping review. PLoS One. 2020;15(3):e0229249.
2. El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. BMC Health Serv Res. 2019;19(1):692.
3. Martínez García L, Pardo-Hernandez H, Superchi C, Niño de Guzman E, Ballesteros M, Ibargoyen Roteta N, et al. Methodological systematic review identifies major limitations in prioritization processes for updating. J Clin Epidemiol. 2017;86:11-24.
4. NIHR Global Research Health Unit on Global Surgery. Delphi prioritization and development of global surgery guidelines for the prevention of surgical-site infection. Br J Surg. 2020;107(8):970-7.
5. Ferguson M, Medley A, Rittenbach K, Brothers TD, Strike C, Ng J, et al. Priority setting for Canadian Take-Home Naloxone best practice guideline development: an adapted online Delphi method. Harm Reduct J. 2022;19(1):71.
6. Wiercioch W, Nieuwlaat R, Zhang Y, Alonso-Coello P, Dahm P, Iorio A, et al. New methods facilitated the process of prioritizing questions and health outcomes in guideline development. J Clin Epidemiol. 2022;143:91-104.
7. Zaror C, Deana NF, Espinoza-Espinoza G, Aravena-Rivas Y, Muñoz-Millán P, Pineda P, et al. Questions and health outcomes prioritization for the development of a COVID-19 dental clinical practice guideline: A case study. J Eval Clin Pract. 2022;28(3):404-10.
8. NIHR Global Research Health Unit on Global Surgery. Global guidelines for emergency general surgery: systematic review and Delphi prioritization process. BJS Open. 2022;6(1).
9. Lynch EA, Lassig C, Turner T, Churilov L, Hill K, Shrubsole K. Prioritizing guideline recommendations for implementation: a systematic, consumer-inclusive process with a case study using the Australian Clinical Guidelines for Stroke Management. Health Res Policy Syst. 2021;19(1):85.
10. Becker M, Neugebauer EA, Eikermann M. Partial updating of clinical practice guidelines often makes more sense than full updating: a systematic review on methods and the development of an updating procedure. J Clin Epidemiol. 2014;67(1):33-45.
11. Sanabria AJ, Pardo-Hernandez H, Ballesteros M, Canelo-Aybar C, McFarlane E, Niño de Guzman E, et al. The UpPriority tool was developed to guide the prioritization of clinical guideline questions for updating. J Clin Epidemiol. 2020;126:80-92.
12. Becker M, Jaschinski T, Eikermann M, Mathes T, Bühn S, Koppert W, et al. A systematic decision-making process on the need for updating clinical practice guidelines proved to be feasible in a pilot study. J Clin Epidemiol. 2018;96:101-9.
13. Sanabria AJ, Alonso-Coello P, McFarlane E, Niño de Guzman E, Roqué M, Martínez García L. The UpPriority tool supported prioritization processes for

updating clinical guideline questions. J Clin Epidemiol. 2021;139:149-59.
14. Sereda M, McFerran D, Axon E, Baguley DM, Hall DA, Potgieter I, et al. A process for prioritising systematic reviews in tinnitus. Int J Audiol. 2020;59(8):640-6.
15. Smith WD. Reweighted range voting – new multiwinner voting method. 2005.
16. Central Committee on Research Involving Human Subjects. Non-WMO research. Available from: https://english.ccmo.nl/investigators/additional-requirements-for-certain-types-of-research/non-wmo-research
17. Central Committee on Research Involving Human Subjects. Your research: Is it subject to the WMO or not? Available from: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not
18. Kok J, Smith WD. Re-weigted Range Voting - a Proportional Representation voting method that deels like range voting. Available from: https://www.rangevoting.org/RRV.html
19. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qualitative Research in Sport, Exercise and Health. 2017;9(4):514-31.
20. Agbassi C, Messersmith H, McNair S, Brouwers M. Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. J Clin Epidemiol. 2014;67(12):1335-42.
21. Sibbald SL, Singer PA, Upshur R, Martin DK. Priority setting: what constitutes success? A conceptual framework for successful priority setting. BMC Health Serv Res. 2009;9:43.

CHAPTER 6
Additional files

SUPPORTING INFORMATION

Additional File 1. Measurement protocol of the think aloud sessions and semi-structured interviews. *This file shows the protocol for the tasks followed during the think aloud sessions and the questions asked during the semi-structured interviews.*

Additional File 2. Published priority-setting assessments mapped over our frame of reference. *This file shows priority-setting assessments mapped to process, function, and outcome components as described by our frame of reference.*

Additional File 3. Qualitative analysis of the usability testing including the resulting action for improvement. *This file shows quotes (Dutch) from the transcribed think aloud sessions and semi-structured interviews enriched with text describing the performed actions on-screen (in brackets). Quotes qualitatively received labels/observations (Dutch with English translation) and were clustered under a generalized interpretation of these labels/observations. Finally, actions for improving the tool's usability were added for improvement of the tool.*

Additional File 4. The REPS-tool programmed in Microsoft Excel. *This file is the complete REPS-tool as a function, where input data can be entered in the function and the semi-automatic ranking within the tool results in output. Please refer to the quick-start guide for information about how to use the REPS-tool in Microsoft Excel.*

Additional File 5. RE-Weighted Priority-Setting: A quick-start guide to the REPS-tool. *This file is a guide to understand and use the REPS-tool in Microsoft Excel.*

Additional File 6. The priority-setting assessment and its outcomes in the Kleefstra guideline pilot. *This file contains figures depicting the priority-setting assessment in the Kleefstra guideline pilot mapped to our frame of reference and its final outcome. Furthermore, the output generated by the REPS-tool under four conditions are shown and compared.*

Additional File 7. The scores of all individuals participating in the Kleefstra priority-setting assessment. *The anonymized file displays all scores the individual participants assigned to key questions in the priority-setting assessment of the Kleefstra guideline. This file is a real-world example that complies to the REPS-tool input format. The participants and their scores can be copy-pasted into the REPS-tool.*

ADDITIONAL FILE 1

MEASUREMENT PROTOCOL FOR THE THINK ALOUD SESSION AND SEMI-STRUCTURED INTERVIEW

1. PREPARATIONS

Email the participant the following documents prior to the session

- PRIORITIZATIONTOOL_EMPTY.XLSX
- SCENARIO_1.XLSX
- SCENARIO_2.XLSX
- FOCUS POINTS.DOCX
- AI_IMAGE3_MOERB.JPG

2. ASK FOR CHARACTERISICS

- Sex (m/f)
- Age (years)
- Overall working experience (years)
- Working experience in guideline development (years)
- Education (Msc./PhD)
- Experience with priority-setting (y/n)
- Have you seen the current priority-setting tool somewhere before this session? (y/n)

3. INSTRUCTIONS (UP TO 5 MINUTES)

[MAIL: FOCUS POINTS.DOCX]

This is a 'think aloud'-session. Here, the purpose is to verbalize your thoughts while performing a task. The goal of the session is to research the usability and completeness of the priority-setting tool for future users. The tool is developed to allow for a ranked priority-setting. Outdated guideline modules are eligible for prioritization.

The session will consist of a warming-up based on a picture, a first introduction to the tool, the use of the tool in two scenarios, and finally a short semi-structured interview. During the two scenarios you will be receiving some specific tasks, however, don't wait for these tasks to start verbalizing your thoughts.

Later on, I would like to ask you to share your screen so I can see how the priority-setting tool is being used. The session is being recorded to produce a transcription. The transcription will be qualitatively analyzed and all data will be anonymized in the data-analysis and research report. You can turn off your camera if you don't want any video recordings of your persona being made.

There are some specific focus points in this think aloud session where I would like to ask some of your attention:

- Stay focused on your task.
- Try to verbalize your thoughts instead of describing your actual actions.
- You do not have to explain or substantiate your thoughts but keep verbalizing your

thoughts.
- Ask any question out loud but the session leader will not answer any question.

Do you have any questions?

You are allowed to revisit these focus points during the session. Refer to the mail I had sent you (focus points.docx).

Are you ready to continue to the warming-up? I will then start the recording.

4. WARM-UP SESSION (UP TO 5 MINUTES)

[TURN ON RECORDING]
[SHARE SCREEN]

I will show you an image. Think of the focus points we just have discussed. Do you want me to repeat them before you start?

[SHOW IMAGE AI_IMAGE3_MOERB.JPG]

Analyze the image and use the think aloud-method to verbalize your thoughts.

[STOP SCREEN SHARE]

Note: The image was created using an artificial intelligence at Artbreeder.com under the creative commons CC0 license.

The image is available through the following link: https://www.artbreeder.com/i?k=2c1196ef89bc996dfef0

5. FAMILIARIZATION WITH THE TOOL (UP TO 5 MINUTES)

[MAIL: PRIORITYZATIONTOOL_EMPTY.XLSX]

Would you start to share your screen from this moment on?

[PARTICIPANT SHARES SCREEN]

Open the tool in Excel. The formulas underneath the spreadsheets are locked. Thus, you cannot break the tool.

1. Look at the tool and browse the tabs. Verbalize your thoughts.
2. Participants score from 0 to 10. Here, 0 = no priority, 1 = the lowest possible priority, and 10 = the highest possible priority. Randomly fill out some scores in the designated cells and see what the tool does.

6. SCENARIO TESTS (UP TO 30 MINUTES)

[MAIL: SCENARIO_1.XLSX]
[CHECK: MAX SCORE = 10, WEIGHT = 0.5, PENALTY METHOD = 0]
[CHECK: NO VALUES IN CELLS CONCERNING RANKING, SCORE, OR ORGANIZATION]

Scenario 1:
Open the Excel-file ‘Scenario_1.xlsx’. This is a file that contains priority scores assigned by participants.

1. Try to get the item names (see tab ‘modules list’) and scores from the scenario Excel-file into the priority-setting tool.
2. Try to rank the scored items by inserting their rank, as a number, above the winner.
3. Try some analyses concerning the central tendency and dispersion measures to discover some modules with a high variance in their scores.
4. See if you can get the ranked list with items from the tool into a Word-file for a hypothetical feedback moment with the working group.

Note: Scenario 1 is a straightforward scenario with a tied ranking once (multiple winners). Scores were randomly assigned in Excel using the =RANDBETWEEN(0;10) function. Only two item scores were manually adjusted to set up a tied ranking for position two.

Scenario 2:
Open the Excel-file 'SCENARIO_2.XLSX'. Participants voted along the interests of their association in this scenario. That is, modules of their 'own' association were assigned high priority scores and other modules were assigned no score or a low priority score. We can apply weights by activating a penalty method. The weight is adjustable in one of the penalty methods.

1. Try to get the item names (see tab 'modules list') and scores in the Excel-file into the priority-setting tool.
2. Try to rank the scored items in the priority-setting tool to create a top-15 by inserting their rank, as a number, above the winner.
3. See if you can get the ranked list with items from the tool into a Word-file for a hypothetical feedback moment with the working group.
4. Fill out the number of voters per association in the priority-setting tool. See the tabs 'Voters' in the priority-setting tool and the tab 'Participation' in SCENARIO_2.XLSX.
5. Clear your top-15 ranking in the row where you just had inserted the ranks. Choose a penalty method in the priority-setting tool and create a new top 15.
   - [If the participant chooses method '2': insert '4' in the cell concerning aggression]
6. See if you can get the ranked list with items from the tool into a Word-file for a hypothetical feedback moment with the working group.

You can stop sharing your screen with me.
[STOP SHARE PARTICIPANT'S SCREEN]

Note: Scenario 2 is a scenario where associations had different numbers of representatives to vote, and those delegates specifically voted along the interest of their 'own' association. Scores were manipulated to achieve this situation. For items of their own association the Excel-function =RANDBETWEEN(8;10) was used while =RANDBETWEEN(0;4) was used for all other modules. Some participating associations did not have 'own' items, the scores of their representatives were randomly assigned using =RANDBETWEEN(0;4).

7. SEMI-STRUCTURED INTERVIEW (UP TO 15 MINUTES)

We will now continue with a semi-structured interview. There are 10 questions where you can provide your opinion about the tool.

- What did you think of the ranked outcome in scenario 2 where delegates voted along the interests of their association?
- What do you think of the current features in the priority-setting tool?
- Which features do you miss when you want to use the priority-setting tool in a priority-setting meeting with a working group?
- What do you think of the overview and structure of the priority-setting tool?
- How could the overview and structure of the tool be further improved?
- What do you think of the ease of use of the priority-setting tool?
- In what areas can the ease of use be improved?
- What should you further know or be able to do before you can use this tool in a priority-setting meeting with a working group?

- What should a working group know or be able to do before this tool can be used in a priority-setting meeting with a working group?
- Do you have any thoughts about the priority-setting tool which have not been addressed until now?

8. CLOSURE

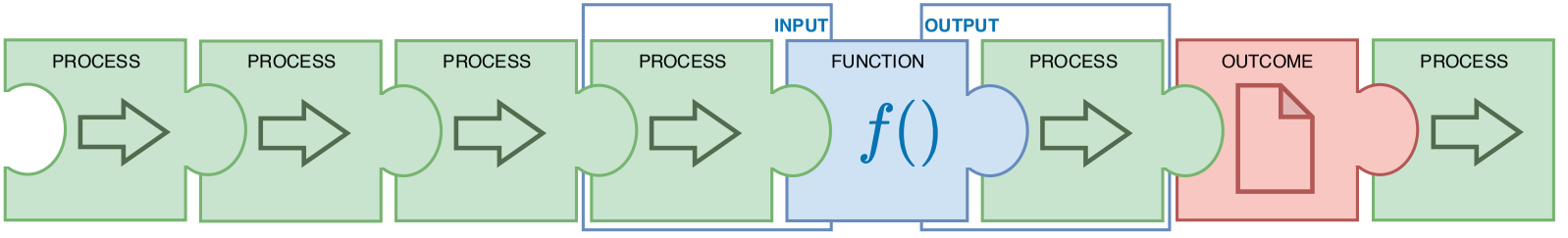
I will now stop the recording
[STOP RECORDING]

The session has ended. We will analyze your input and with that we hope to improve the priority-setting tool. Thank you very much for your participation!
[CLOSE VIDEO CONFERENCE CALL]

NB1: Two penalty methods were present in this preliminary version of the tool. In later versions we removed one of the methods, which also expired the need for using the tab 'voters' in the tool.
NB2: The remaining penalty-adjusted weighting method and penalty aggression parameter were later renamed to decay-adjusted weighting method and decay aggression, respectively.

ADDITIONAL FILE 2

MAPPING PUBLISHED PRIORITY-SETTING ASSESSMENTS TO THE FRAME OF REFERENCE



Establishing a working group

The original guideline developers were contacted for participation. A person with a similar profile was contacted when the original guideline developer was not available for participation. Conflicts of interest were collected from all working group members. Training about the UpPriority tool was provided.

Mapping clinical questions

Clinical questions were identified in the original clinical practice guidelines and were linked to the relevant guideline sections and recommendations.

Survey development

An online survey was developed to assess the clinical questions. It also captured the participants' background and feedback on the UpPriority tool. Weekly reminders were send to participants up until five weeks, whereafter the survey was no longer available. Support was provided to participants upon request.

Priority assessment of the clinical questions

All of the participants assessed all of the clinical questions within a clinical practice guideline according to six criteria (i.e. safety, evidence, context, method, user, access) on a 7-point Likert scale. A higher score meant a higher priority for updating.

Calculation and ranking

Input: The participant's individual scores on all individual criteria for each of the assessed clinical questions.

Function: An overall mean score and its standard deviation for each of the clinical questions was calculated based on the scores on each priority item given by the participants. Clinical questions were ranked high to low on priority score and thereafter on standard deviation (low to high).

Output: A list of ranked clinical questions

Updating decision

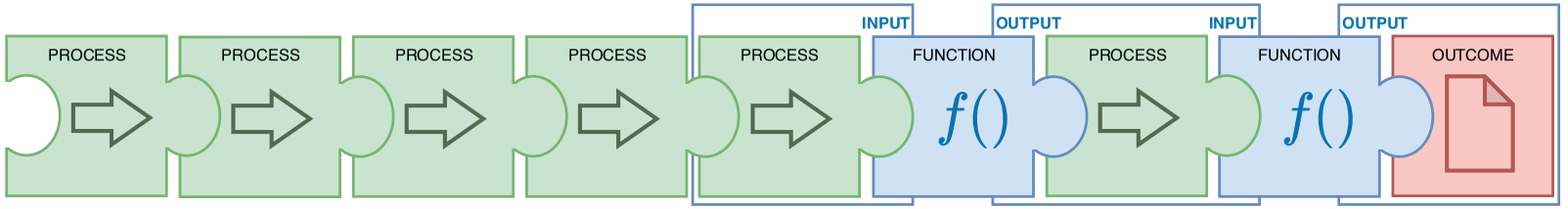
The ranked list of clinical questions was used to classify the clinical questions into three groups of priority: high priority for updating, medium priority for updating, or low priority for updating. The classification was further informed by an alert threshold (i.e. priority score ≥ 30 or item score ≥ 5) and additional considerations or advice from the working group's participants.

The updating decision resulted in a definitive list of clinical questions classified in low, medium, or high priority for updating.

Writing a report

A priority report was written, containing: a summary, the objective, the methods used, a list of the clinical questions, the ranking of the clinical questions, the priority decision, and reasons for the priority decision. All working group participants approved the priority report.

A2 Figure 1. The priority-setting assessment described in Sanabria (2021) was mapped to the components in the frame of reference. *Visualized by M.S. Oerbekke using the frame of reference. Assessment from: Sanabria AJ, Alonso-Coello P, McFarlane E, Niño de Guzman E, Roqué M, Martínez García L; UpPriority Implementation Working Group. The UpPriority tool supported prioritization processes for updating clinical guideline questions. J Clin Epidemiol. 2021 Nov;139:149-159. doi: 10.1016/j.jclinepi.2021.07.022. Epub 2021 Aug 5. PMID: 34363971.*



Initial key questions

Key questions were identified from a published scoping review

Key question review

Three committees reviewed the key questions for content, completeness, and readability. New key questions could be added.

Key question translation

Key questions were converted to plain language questions.

Online survey development

Data collection tools were developed and piloted for online administering using REDCap.

Assigning importance 1

The guideline panel was asked to rate 35 questions on a 5-point Likert-scale. Here, panel members were asked whether they agreed that the question is important to discuss in a best-practice guideline (1: strongly agree, 2: agree, 3: neutral, 4: disagree, 5: strongly disagree) and to provide a rationale for their rating.

Delphi round 1

Input: Guideline panel ratings for each member for 35 questions.

Function: Mean scores were calculated. Mean scores below 2.0 were included for the next round. Mean scores of 2.0 and above were excluded.

Output: A list of 17 questions that scored below 2.0.

Assigning importance 2

Panel members were asked to consider reasons for (dis)agreement of other panel members and to rate the remaining 17 questions again.

Delphi round 2

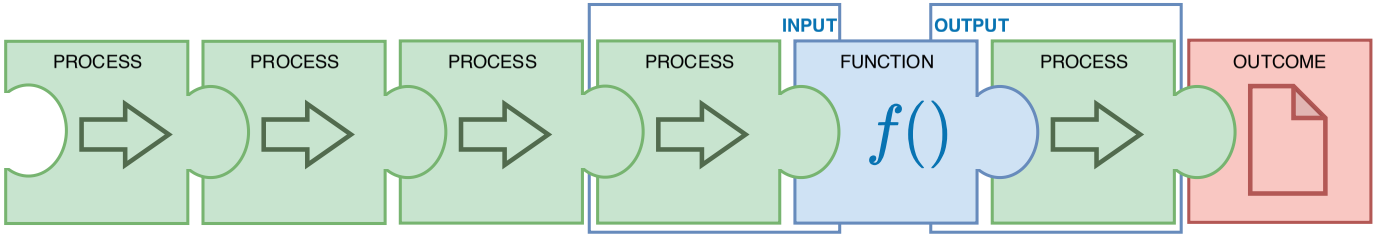
Input: Guideline panel ratings for each member for 17 questions.

Function: Mean scores were calculated. Mean scores below 2.0 were included on the list. Mean scores of 2.0 and above were excluded.

Output: A list of 9 questions that scored below 2.0.

The second Delphi round resulted in a final list of 9 questions prioritized for the take-home naloxone guideline

A2 Figure 2 – The priority-setting assessment described in Ferguson (2022) was mapped to the components in the frame of reference. *Visualized by M.S. Oerbekke using the frame of reference. Assessment from: Ferguson M, Medley A, Rittenbach K, Brothers TD, Strike C, Ng J, Leece P, Elton-Marshall T, Ali F, Lorenzetti DL, Buxton JA. Priority setting for Canadian Take-Home Naloxone best practice guideline development: an adapted online Delphi method. Harm Reduct J. 2022 Jul 2;19(1):71. doi: 10.1186/s12954-022-00650-4. PMID: 35780136; PMCID: PMC9250272.*



Brainstorming

Panel members from 10 guidelines brainstormed to generate a list of potential questions for their guidelines

Priority criteria selection

Six criteria were selected from literature and guideline methodologist input. Questions would be rated on whether it was one:
- *That commonly arises in practice*
- *For which there is uncertainty in practice regarding how to manage patients*
- *For which there is new research evidence to consider*
- *That is associated with variation in practice*
- *That has important consequences for, or is associated with, high resource use or costs*
- *That has not been previously or sufficiently addressed (e.g. in previous guidelines)*

Survey development

Questions were formatted using the PICO acronym. The survey asked panel members to rate the questions from one (least important) to nine (most important) based on the six criteria.

Priority assignment

All panel members were asked to provide the overall rating of importance, based on the six criteria.

(NB: participants were randomized for the first question of the survey whether to provide an overall score or to provide scores for all six criteria and thereafter an overall score. Starting from the second question, all participants rated six criteria and provided an overall importance score).

Priority summary

Input: The participant's overall importance scores based on all six criteria for each of the assessed questions.

Function: Panel members' score were summarized as mean ratings per item. Means were rounded to the nearest full digit. The mean rating was color coded and categorized as high priority (mean score 7-9, green), important but not high priority (mean score 4-6, yellow), or low priority (mean score 1-3, red).

Output: A list of questions. Out of 469 questions, 119 received a high priority to address in the guidelines, 340 questions received important but not high priority, and 10 received low priority.

Discussion

Priority results were presented to the panel members and used to guide discussions between panel members to reach consensus on a final list of questions. The first dicision was based on the overall rating. Ratings on the additional criteria were used to choose between question with the same/similar overall importance scores.

The discussion resulted in a final list of high priority questions to be addressed in guidelines regarding venous thromboembolism.

A2 Figure 3. The priority-setting assessment described in Wiercioch (2022) was mapped to the components in the frame of reference. *Visualized by M.S. Oerbekke using the frame of reference. Assessment from: Wiercioch W, Nieuwlaat R, Zhang Y, Alonso-Coello P, Dahm P, Iorio A, Manja V, Mustafa RA, Neumann I, Ortel TL, Rochwerg B, Santesso N, Vesely SK, Akl EA, Schünemann HJ. New methods facilitated the process of prioritizing questions and health outcomes in guideline development. J Clin Epidemiol. 2022 Mar;143:91-104. doi: 10.1016/j.jclinepi.2021.11.031. Epub 2021 Nov 26. PMID: 3484386*

ADDITIONAL FILE 3

QUALITATIVE ANALYSIS OF THE USABILITY TESTING

A3 Table 1 - Part 1: Familiarization

P	Participant text	Label/observation	Interpretation	Actions to improve
1	Er staan heel veel items zo te zien. Het gaat door tot 100. Ehm, een aantal formules zie ik staan. Ik kan nergens op klikken [klikt willekeurig cellen aan op het tabblad RRV], ik weet niet of dat klopt… alleen de tabbladen.	1e indruk: veel items [First impression: lots of items]	First impression: large and complex	Provide an introduction to the tool's structure in an accompanying document
1	[klikt tabblad RRV aan] Ja, het eerste tabblad vind ik overweldigend [scrollt horizontaal], ziet er heel ingewikkeld uit. [scrollt verticaal]	1e indruk: overweldigend [First impression: overwhelming]		
2	Oke, ehmm. Op het eerste ook zie ik een best wel complex-ogend.. systeem. Ehm… Waar soort van optelsom van iets van lijkt te worden gemaakt.	1e indruk: complex ogend [First impression: looks complex]		
3	Het eerste wat mij opvalt is dat hier allemaal 'hash-tag getal uitroepteken' staat. Dat zijn natuurlijk formules die er onder liggen. Tenminste dat eh, dat lijkt mij zo. Het ziet er wel een beetje daunting uit [scrolt naar boven en beneden], dat ik denk: oeh, daar moet ik iets invullen! Maar dat is vast niet zo.	1e indruk: afschrikwekkend [First impression: daunting]		
4	[[scrolt door tabblad 'RRV'] Ja, nou, eerste wat ik opende dacht ik: wow wat is dit groot, ofzo.	1e indruk: groot [First impression: large]		
2	Nou de eerste feed is zeg m… Of de eerste tabblad is mij zeg maar niet per se duidelijk wat er nu moet komen te staan, behalve dat hier 'winner' [wijst met cursor naar rij 3 'Winner'] staat. (…) klikt op tabblad labels list, scrollt verticaal] Dit is gewoon allemaal leeg. Geen idee wat hier allemaal ingevuld moet worden, misschien de.. zo de individuele modules. (…) [klikt op tabblad verenigingen] Nou, hier nog een lijstje met de verenigingen… [scrollt verticaal] met de afkortingen er achter. Wat was dit dan? [klikt op tabblad labels list] Even kijken, hoor. [klikt op tabblad voters, scrollt verticaal] Oh daar kan je het waarschijnlijk, 't is een soort van tabblad wat gewoon, [klikt op tabblad verenigingen] alleen de afkorting geeft. [klikt op tabblad RRV] Nog een keer naar het eerste tabblad. Dit is echt eindig, oneindig lang. Ik weet niet wat RRV ranking betekent en ik snap het ook niet.	Structuur van de tool niet direct helder [Structure of the tool is not immediately clear]	Structure may be unclear at first	Provide an introduction to the tool's structure in an accompanying document

	(…) [scrollt verticaal] Nou, ik denk dat dit uiteindelijk een soort van de overzichtspagina moet zijn waarop de scores straks komen te zien. Maar ik snap nog niet zo goed hoe dat precies moet gebeuren.			
4	En dat ik nu even niet het idee heb wat moet ik nu hier mee doen. Ik heb echt geen idee wat de bedoeling hier bij is, als ik het zo zie. (…) En als ik vooral naar de eerste pagina [klikt op tabblad 'rrv'], of het eerste tabblad, die lijkt mij het belangrijkste en dank ik… dan heb ik echt [scrolt naar beneden] even geen idee wat ik daar… in moet vullen. (…) Oh nee. Maar ik heb dus inderdaad geen idee hoe je dit in zou moeten vullen, nu zeg maar.	Structuur van de tool niet direct helder [Structure of the tool is not immediately clear]		
3	Ahja, dit is, dit is wel handig [wijst met cursor naar uitleg bij tool parameters]. Hier staan bepaalde scores, en ook al wat ingevuld met ook een uitleg erbij. Dat is prettig.	Ervaart de begeleidende teksten in de tool als prettig [Experiences the guiding texts within the tool as helpful]	Helpful guiding texts	-
3	Ik krijg een klein beetje.. eh.. associaties met, eh, SPSS, waar je ook in verschillende tabbladen dan verschillende, eh, typen informatie kan invullen.	Vergelijking met andere software [Comparing to other software]	Initial similarities	-
3	[klikt op tabblad 'Ranking outcome'] (…) Dus dit is, dit is eigenlijk om het.. gebruikersgemak nog verder te vergroten. Namelijk, om een soort tool aan te bieden dat je de eh, de.. de ranking van de modules eh, gewoon ergens anders naartoe kan kopiëren.	Ziet hoe de manier van output is bedacht en ervaart dit als het vergroten van gebruiksgemak [See show the output was designed and thinks this inreases usability]	Initial thoughts on tool output	-

A3 Table 2 - Part 2: Adding scores

P	Additional instruction	Participant text	Label/observation	Interpretation	Actions to improve
1		[klikt op vergrendelde cellen] Maar waar kan ik dan klikken? Hier [selecteert cel waar scores ingevuld kunnen worden]? Bij scores? (…) [selecteert cel waar de rank kan worden aangegeven] Oh wacht, hier kan ik ook wat doen. En de bedoeling was dat ik scores ging invullen nu, he? [vult 0 in de cel in] (…) Van de deelnemers…. Maar ben ik dan zelf de deelnemer? [klikt tabblad Voters aan] Of moet ik hier… [klikt tabblad RRV aan] Nee… [selecteert cellen voor de rank en vult er willekeurig cijfers in] Ik neem aan dat ik hier dan wat moet invullen. Hoeveel wil je dat ik er doe?	Structuur niet helder: items, deelnemers, scores [Unclear structure: placement of item-names, participants, scores]	Unclear placement of item names, participants, and scores at first	Provide an introduction to the tool's structure in an accompanying document
2		Ja. Even kijken, dan denk ik [klikt op tabblad ranking outcome, klikt op tabblad labelslist] dat we in.. Hmm.. Zal dat in deze moeten? Ik ga gewoon even proberen. [vult cijfers in de cellen die bestemd zijn voor de itemnamen in] Acht, Negen, drie, twee, vier.	Niet direct duidelijk waar de scores ingevuld moeten worden [Unclear structure: placement of scores]		
2		He?! Even kijken hoor. [selecteert de cel voor module 1 die bedoeld is voor de ranking] Oke, zeg dat ik een 10 geef aan de eerste module [vult 10 in de cel in]. Oke wacht, en een 7, [vult scores in voor modules 2 t/m6 in de cellen die bestemd zijn voor de rank] en een 8, en een 3, en een 2, en een 5. Hmm, even kijken wat er nu gebeurd. [scrollt verticaal] Ik zie eigenlijk nog… [kucht] Maximale score… [inaudible] Ehm, ik vraag me af of ik het wel goed doe, want… [klikt op tabblad voters] Of ik nou de opdracht wel begrepen heb?	Niet direct duidelijk waar de scores ingevuld moeten worden [Unclear structure: placement of scores]		
4		Dus als ik zeg: hier 2. [klikt de correcte cel aan voor de eerste persoon in de kolom voor namen] En hier dan even een naam noemen, [vult in: 'module heup'] module heup.	Structuur: plaats scores en item-namen niet direct helder [Unclear structure: placement of scores and item names]		

4		Maar als ik dan hier [klikt de cel voor organization aan voor de tweede persoon] nog eentje invul, bijvoorbeeld <*wetenschappelijke vereniging*> [kopieert en plakt de cel van de eerste persoon voor de tweede persoon], en dan hier weer module heup [kopieert 'module heup' in kolom naam van de eerste persoon naar de cel voor naam voor de tweede persoon] , wat gebeurt er dan? Krijg je dan twee mensen die…? Oh ja, twee mensen die het gescoord hebben. Totale… en het gemiddelde is dan 2. Ahh, oke. En als ik dan hier een andere module doe van de.. [selecteert cel voor de organisatie van de derde persoon, vult in 'NOV ' en krijgt een melding dat deze organisatie niet voorkomt in de lijst] ho, nja, oh dan kan je natuurlijk een keuze maken [opent het drop-down menu] uit het tabblad die daar staat. Nou, als ik dan hier ook de <*wetenschappelijke vereniging*> doe… [selecteert <*wetenschappelijke vereniging*> uit het drop-down menu] en dan hier [selecteert de cel voor name van de derde persoon, vult in: 'module knie'] 'module knie'.. [selecteert de cel voor de score van persoon drie op item 1, vult 3 in] dan zou je hier score drie krijgen. Hey? Ik had gedacht dat je dan hier weer een nieuwe rij op rij G of wat is het? (…) modules moeten denk ik hier [klikt op tabblad 'ranking outcome']. Maar hier krijg ik dan niks [scrolt op en neer]. Je zou denken dat je dan hier een soort score uit krijgt [klikt op tabblad 'verenigingen', klikt op tabblad 'ranking outcome', klikt op tabblad 'RRV']. Maar ik snap ook überhaupt niet waarom je dan hier [wijst met cursor naar item namen in rij 1] nu pas de module knie ziet staan. En niet als ik hier [wijst met cursor naar de kolom voor namen van deelnemers] dan bij 'naam' module knie neer zet, dat het dan allemaal bij elkaar komt in deze rij.	Structuur: verwisseling van naam deelnemer met naam item in de tool [Unclear structure: placement of participants and itemnames]		
1	[I]: Hmm-Hmm. Zou je die rij eh eens voor mij helemaal leeg willen maken? Die rank? [P01]: Die rank helemaal leeg? [I]: Ja, ja wat je net hebt ingevuld? Oke, en bij de cellen bij scores… zie je 'm staan? Daar, ja. Zou je daar eens getallen willen invullen?	[maakt de rij voor ranking leeg en selecteert een cel waar scores ingevuld kunnen worden] Hmm-Hmm. Even kijken, dan ga ik hier dingen invullen [vult voor item 1 meerdere willekeurige scores in]. Dan krijgen we een winner al te zien. [lacht]. Oh wacht, net heb ik de rank ingevuld en nu is.. pas de scores. Dan zat ik net denk ik verkeerd. (…) Maar nu staat er alleen bij item 1 wat, voor de rest nog niks. [vult één score in voor zowel item 2 als 3] Ah zie, dus dat doe je onder elk item. Oh dus nu heb ik voor verschillende organisaties wat ingevuld, eigenlijk voor verschillende mensen [selecteert lege cellen waar naam ingevuld kan worden]. Dus nu heb je meerdere voters voor 1, als ik het goed zie…	Structuur wordt helderder na een kleine aanwijzing: rank, score, deelnemers, items [Structure becomes more clear after a small instruction: placement of rank, scores, participants, item names]	Placement of scores, participants, itemnames, and ranks becomes more clear after some instructions	Instruction and background information seem to be helpful to recognize the tool's structure. Provide an introduction to the tool's structure in an accompanying document

2	[I]: Oke, je mag de.. de rank wat je nu net hebt ingevuld in die rij… [P02]: Deze? [selecteert de cel met de rank voor module 1 in rij 2] [I]: Ja, die mag je even helemaal leeg maken. [P02]: [verwijdert alle ranks in rij 2] [I]: Oke, en onderin, als je ietjes naar beneden scrolt, [P02]: [scrollt verticaal] [I]: ..daar staan de scores. [P02]: heb je het over maximum…? [wijst naar cellen m.b.t. max score en penalty] Waar precies moet ik kijken? [I]: Nee, nog iets naar onder. [P02]: [scrolt verticaal] [I]: Daar heb je organisatie, name, scores… [P02]: Ja.. [klikt de cel aan die bestemd is om scores in te vullen voor de eerste persoon en eerste module] [I]: En bij scores mag je wat invullen als je wil. [P02]: Ohke. [vult 8, 9, 1, 2, 4 in voor 5 verschillende personen voor module 1] Zoiets? [I]: Ja.	Ja? [scrollt verticaal] Hey… Hmm? Oke, wacht even. Nu ineens is zeg maar de eerste module winnaar geworden omdat ‘ie 24 punten heeft. [scrollt verticaal] Ik snap alleen nu niet hoe dit nou kan, dat er… [scrollt verticaal] Waarom ‘ie dat nou alleen op de eerste module.. Ohh, wacht, omdat ‘ie in de kolom van 1 zit waarschijnlijk. Als ik het hier doe… [vult cijfers in voor 5 verschillende personen voor module 2] vijf, vijf, vier, drie, twee. [scrollt verticaal] Kijk, dan krijgt ‘ie ook een score. [scrollt verticaal] Ahh, oke. Ohh, oke.	Structuur voor scores invullen wordt duidelijker na een aanwijzing [Structure becomes more clear after a small instruction: placement of scores]		

2		Ah wacht, oke. [klikt op tabblad labels list] Dus dit is waarschijnlijk alleen maar een lijst waarin je… hier [selecteert de cel voor de eerste modulenaam] moet je waarschijnlijk je module zo neerzetten. [voert de tekst 'module 1' in de cel in, klikt op tabblad RRV]] En dan, ja.. Dit is… oke. Dus dit is alleen de titel.	Structuur: link tussen tab labels list en item namen in RRV tab is duidelijk [Structure: link between labels list tab and RRV tab is clear]	Placement of item names, participants, organizations, and/or scores may be clear to some without prior instructions	-
3		Even kijken [scrollt naar beneden]. Ohja, hier kan ik dus de organisatie invullen [opent drop-down menu met alle organisaties] met een drop-down menu [scrollt door het drop-down menu]. Bijvoorbeeld de <*wetenschappelijke vereniging 1*> [klikt <*wetenschappelijke vereniging 1*> aan in het drop-down menu]. [klikt op cel D35 in de kolom 'name'] En dit is dan de naam van degene die de score maakt, dus dat is, eh, dat ben ikzelf [vult haar eigen naam in]. En.. [klikt op het drop-down menu voor organisatie op rij 36] De <*wetenschappelijke vereniging 2*> [selecteert<*wetenschappelijke vereniging 2*> uit het drop-down menu]. Eh, noem eens wat. <naam persoon> [vult de naam van deze persoon in op de correcte plaats]. [selecteert NVDV in het drop-down menu in rij 37, vult een naam in in rij 37 op de correcte plaats]	Ziet zelf direct de structuur waar deelnemers en organisaties moeten worden ingevuld [Immediately sees the structure: placement of participants and organizations]		
3		[klikt op tabblad 'Labels list'] Dat was deze. Even kijken [vult namen van modules in op de juiste plaats]. Ik doe het gewoon even makkelijk, want [inaudible]. Drie, vier, vijf.	Ziet zelf de structuur waar de itemnamen moeten worden geplaatst [Sees the structure: placement of item names]		
3		Nou, ik vind: drie, vier, een, acht, negen [vult deze score in op rij 35 op de juiste plaats]. De laatste vind ik dan heel belangrijk. <naam> vindt dan juist eh, deze en deze heel belangrijk [vult scores in op rij 36 op de juiste plaats]. Die medium, deze dan weer laag. En <naam> vindt eigenlijk alleen maar deze heel belangrijk [vult scores in op rij 37 op de juiste plaats]. Tien zelfs.	Ziet zelf de structuur waar de scores moeten worden geplaatst [Sees the structure: placement of scores]		
3		Oh kijk, nou komt er meteen al een 'winner' uit. Module 4. [scrolt naar beneden] Het lijkt er op, ja er zijn natuurlijk twee mensen die hem hebben geprioriteerd dus dat is ook wel logisch.	Ziet zelf direct dat de tool automatisch een winnaar identificeert [Immediately sees that the tool automatically identifies a winner]	Automatic winner identification can be observed without prior instructions	-

3		Want [klikt op tabblad 'ranking outcome'], staat dat dan meteen hier al in? Nee, nog niet [klikt op willekeurige cellen onder de kolommen 'name' en 'rank', klikt op tabblad 'RRV', klikt op tabblad 'ranking outcome']. Nou misschien heb ik het dan toch anders geïnterpreteerd dan ik dacht. Ik kijk in de ranking outcome en dan staat er nog niet iets	Manier van ranken nog onduidelijk: rol toekennen van ranks i.r.t. de ranking uitkomst nog onduidelijk [Way of ranking still unclear: role of assigning ranks related tot he ranking outcome is still unclear]	Unclear role of assigning ranks	Provide information in an accompanying document about how to assign ranks and its role in relation tot he 'ranking outcome' tab.
3		Ik vraag mij af.. Ehm, moet ik dan hier ook nog number of voters invoegen [wijst met cursor cel aan bij number of voter bij <*wetenschappelijke vereniging*>] , aangezien we hier [klikt tabblad 'RRV' aan] ook de naam en organisatie invullen? Zal die dan dan ook niet automatisch kunnen… [klikt tabblad 'Voters' aan] eh… kunnen doorlinken naar deze tab [klikt tabblad RRV aan]?	Wens om invullen van aantal voters te automatiseren [Wish to automate filling out the number of voters]	No automation for number of voters	We have decided that we will use a different weighting method independent of the number of voters per organization. Therefore, the number of voters per organization does not have to be filled out in the next iteration of the tool.
1		Even testen of dat nu ook zo is als ik meer invul [vult meerdere scores in voor items 2, 3, 4]. Even kijken. Wordt gewoon random… Oh 52 kan niet, ik krijg geen foutmelding als ik 52 invul.	Check voor (non) valide scores in tool [Check for (non)valid scores in the tool]	Missing check for valid scores	We will add a button that checks for input larger than the maximum scale score and less than 0 in the score matrix.

A3 Table 3 - Part 3: Using the tool in scenario 1

P	Additional instruction	Participant text	Label/observation	Interpretation	Actions to improve
1		Maar dit [selecteert cellen met modulenamen in de header op tabblad database in scenario 1] is makkelijker kopiëren als het in de itemlijst moet, toch? Of hoe zag dat ding er uit? [klikt op de cellen met itemnamen in het tabblad RRV] Kan hier niks in knippen, plakken… [klikt op tabblad voters, klikt op tabblad labels list] Dat moest natuurlijk hier. Ja. Oke… Dan is deze [klikt op modulenamen in tabblad modulelijst in scenario1] inderdaad makkelijk.	Input: plaatsing itemnamen eenvoudig [Input: easy placement of itemnames]	Getting data into the tool is generally easy	-
2		Even kijken, dan ga ik naar eh [klikt op tabblad labelslist] die, denk ik. Label list. [plakt modulenamen in de correcte cellen] (…) Ik kopieer er nog een van de richtlijn duizeligheid bij ouderen. [plakt modulenamen in de correcte cellen] (…) Oke, ik heb nu in [selecteert alle cellen met module namen] totaal 19 labels gekopieerd, [klikt op tabblad RRV] die staan nu in mijn overzichtspagina ook netjes bovenaan.	Input in tool: eenvoudige input van item namen in de tool [Easy input of item names in the tool]		
2		Eehm, even kijken, ik heb hier eerst de vereniging. Die zet ik bij organizations. [plakt de verenigingen in de juiste cellen] (…) [plakt deelnemersnamen in de correcte cellen] De namen van de mensen er bij. En dan heb ik…. Kijk dit gaat gesmeerd. Hatsjee. [plakt de scores van de deelnemers in de correcte cellen] Ik plak het er in, en als het goed is.. [scrollt verticaal] Ha, kijk! Lichamelijk onderzoek bij duizeligheid is onze winnaar.	Input in tool: eenvoudige input van organisaties, deelnemers en scores in de tool [Easy input of organizations, participants, and scores in the tool]		
3		Ehm, dan kopieer ik eh.. de namen van de modules van de eerste richtlijn. Die zet ik dan in de tool… [klikt op tabblad 'labels list'] bij labels list [selecteert 5 cellen, plakt module titels in de 5 correcte cellen]. Ik doe dan eigenlijk niks met… de namen van de richtlijnen [klikt op tabblad 'RRV]. Volgens mij is dat ook.. niet zo'n probleem, want die kan ik ook niet kwijt verder. Eh, en het is ook niet zo relevant natuurlijk omdat we richtlijnoverstijgend gaan denken. Dus, de modules kopieer ik. Ik ga met het volgende rijtje modules door. Eh.. bij labels list [klikt op tabblad 'labels list'], zo, [plakt volgende set moduletitels aansluitend in tabblad 'labels list'] copy-paste. En de laatste richtlijn. Hier. Zo [plakt volgende set moduletitels aansluitend in tabblad 'labels list']. [klikt op tabblad 'RRV']Even kijken hoe dat er uit ziet. Inderdaad, ik heb nu… de modules [scrolt horizontaal heen en weer], de naam van de modules boven in staan.	Input: manier van itemnaam input naar de tool krijgen gaat eenvoudig [Manner of item name input to the tool is easy]		

3		Met de namen er in te zetten.[scrolt naar beneden, plakt de rij namen in de juiste kolom] En de organisaties die daar bij horen [plakt de organisaties in de juiste kolom]. Yes, dat pakt 'ie ook.	Input: manier van deelnemers en organisaties input naar de tool krijgen gaat eenvoudig [Manner of organization input to the tool is easy]		
3		Oke, dus dan hebben we de scores nog, die zal ik dan ook even kopiëren. Zo. En dan… zet ik die ook [plakt de matrix van scores op de juiste plaats] hierin.	Input: manier van score input naar de tool krijgen gaat eenvoudig [Manner of score input to the tool is easy]		
4		[gaat naar de tool, klikt op tabblad 'labels list'] In deze. [selecteert en kopieert de moduletitels van de eerste richtlijn in het databestand en plakt deze in de tool]	Input: manier van itemnamen input naar de tool krijgen gaat eenvoudig [Easy way of getting item names into the tool]		
4		Ja, ik weet niet [plakt de namen op de juiste plaats in de tool] of ik het nu in een keer weer goed zou doen, dat ik dan zou, zeg maar, nu weet ik dat er dan hier wordt bedoeld dat je dan al die scores zo daar neer zet. [kopieert de scores uit de dataset] Maar ik weet niet of ik dat dan de volgende keer… [plakt de scores op de juiste plaats in de tool] gebaseerd op wat hier dan nu alleen staat ook weer zou doen. Dan zou ik toch hier [wijst met de cursor naar rij 34 boven waar scores ingevuld worden] misschien ook wat neerzetten ofzo. Of dit [klikt op de tekst 'scores' in rij 34] iets aanpassen, waardoor dat dan duidelijker is dat je hier dus los de scores weergeeft.	Plaatsing correct, maar wens meer sturing over de plaats van de input van scores in de tool [Place of score input in the tool is correct, but wishes for more guidance in the tool]		
4		Maar de richtlijn weet ik niet of je die er dan ook bij moet benoemen. Maar ik zou zeggen, zelf zeggen, dat je dit [selecteert en kopieert de moduletitels van de tweede richtlijn in het databestand] hier in zet [plakt de module titels aansluitend in de tool]. Lijkt mij wel handig dat je dan ergens zometeen ook kan zien… [selecteert en kopieert de moduletitels van de laatste richtlijn in het databestand, plakt de moduletitels aansluitend in	Brontitel wordt meegenomen als item in de tool [Title of the item's source is used as an item in the tool]	There is a wish or consideration to identify the source of the items in the tool	We will not adjust the tool, but rather add guidance about source identification in a document accompanying the tool.

		de tool] wat dan de richtlijn is. (…) Want nu zie ik al die namen hier wel staan, maar ik zou ook willen [klikt op tabblad ‘labels list’, klikt op tabblad ‘ranking outcome’] dat je dan kan zien… [klikt op tabblad ‘labels list’] uit welke richtlijn het komt. (…) Stel je hebt toevallig twee richtlijnmodules met precies dezelfde naam en je gaat ‘m prioriteren en.. je komt eruit, dan weet je niet meer uit welke richtlijn het komt. Ho. [klikt op Word ipv Excel] Dus dan zou ik toch zeggen, omdat ik nu niet zo heel goed weet hoe het werkt, om dan toch hier achter even die naam te kopiëren [kopieert de richtlijn naam van de eerste richtlijn in de dataset en plakt deze achter de module titel van het eerste item].			
1		Moet de richtlijn er zelf ook bij? Ik denk het niet. [kopiëert en plakt modulenamen in de cellen in tabblad labels list]	Identificatie van de bron van items [Identification of the source of the items]		
1		Dan moet ik denk ik toch hier zijn. Dan verenigingen [kopieert en plakt verenigingen van scenario1 naar de correcte cellen in de tool], dan namen…[kopieert en plakt de namen van scenario1 naar de correct cellen in de tool] en dan de scores per.. ding [selecteert alle scores in scenario1].	Leereffect na aanwijzing in ‘scores invullen’ (vorige opdracht) [Learning effect after instructions in previous assignment]	When the structure of the tool was unclear, participants were able to correctly place the tool’s input after a little instruction	Provide an introduction to the tool’s structure in an accompanying document
1	[I]: Net.. Net vulde je wat in op die eh.. ranking, die rij. [P01]: Ja [klikt de eerste cell in de ranking rij aan] [I]: Eh, als je nu boven de Winnaar eh.. 1 zet… Ja..	[vult 1 in de cel boven winnaar in] Dan krijg je andere soort [inaudible], maar moet ik dan alsnog…? Oh wacht deze [selecteert de cel boven een MULTI(2)] is dan nu gelijk tweede. Krijgen ze dan allebei een 2? [vult 2 in boven de cellen van MULTI(2)] Oh, hij geeft nu zelf aan wat een… [vult ranking in voor 4 en 5] De winnaar is als je wat invult. [vult rank 6 in] Oke.. Ik heb nu 3 overgeslagen trouwens omdat ik twee keer 2 had, maar… maarja die staat er dan ook. Zes, Zeven [vult ranks in].	Leereffect: Begrijpt de manier van ranken na kleine aanwijzing [learning effect after a little instruction]		
4	I]: oke, ik ga je wat dat betreft… Want dit punt eh, hier zat je net ook op en dus	Ohh. En de organisatie is dan… oke! [selecteert de rij met namen uit de dataset en kopieert deze] De naam van de mensen die scoren.[plakt de lijst met namen op de correcte locatie in de tool]	Herkent de structuur na een kleine aanwijzing en vult de input op de correcte		

	dat wordt wel helder uit de analyse. Ik ga je hier eventjes een klein beetje sturen. De naam, dat zijn de namen van de mensen die scoren.	(…) En dan de vereniging, die moet dan daarbij. [selecteert en kopieert de lijst met verenigingen in de dataset, plakt de lijst met verenigingen op de juiste plaats in de tool] En dan de scores die kan ik dan ook gewoon… [kopieert de scores voor het eerste item uit de database en plakt deze op de juiste plaats in de tool] Oh, wacht, denk ik, ofniet? Nu, hij komt natuurlijk gewoon onder de module denk ik, ahh! [kopieert de scores van item 2 in de dataset] Nu snap ik het denk ik, [plakt de scores voor item 2 op de juiste plaats] want dan zet dit hier onder, jaa! Oh ik moest even een beetje, [lacht][kopieert de scores voor module 3 uit de dataset en plakt deze op de juiste plaats in de tool]. Nah, [kopieert de matrix van scores van de overige items uit de dataset] anders zou er misschien ook wel modules staan ofzo.	plaats in de tool in. [Recognizes the structure after a little instructin and fill out the tool's input in the correct place]		
1		Ik ga er even vanuit dat het op de juiste volgorde staat [plakt de scores in de juiste cellen in de tool]. Als het goed is.. klopt het dan nu [vergroot de tool naar full screen]. Ja, volgens mij is het nu compleet.	Check volgorde items en volgorde scores [Checking the order of items and scores]	Participants seem to understand that the order of the scores must align to the order of the items.	Although it seems to be generally clear, it is an important aspect. We will elaborate about the input format in an accompanying document to the tool.
3		Even kijken, posterieur kanaal diagnose. Dat komt overeen met wat onder de richtlijn benigne paroxysmale positieduizeligheid valt. Ja, het zijn er heel veel, en de volgorde komt volgens mij ook overeen.	Check volgorde van lijst met itemnamen en scores of deze overeenkomen [Checks the order of the item names and scores]		
3		Ik zou eigenlijk moeten checken of alle modules, of de volgorde van alle modules helemnaal overeenkomt… met de volgorde van de modules.. die ik net in dat lijstje heb gezien. Nja, voor de snelheid ga ik er maar even vanuit dat het klopt.	Check volgorde van lijst met itemnamen en scores of deze overeenkomen [Checks the order of the item names and scores]		
4		Klopt het dat dat dan van, dat die andere modules niet gescoord zijn? Dus tot: ziekte geassocieerd. [controleert of alle items in rij 2 een score hebben in de kolom] Maar dat komt dan niet. Oh, dat komt bij mij natuurlijk niet overeen omdat ik	Controle of score-matrix overeenkomt met de itemlijst; komt niet overeen		

		richtlijn…, ahh, daarin heb genoemd.	omdat de brontitels zijn meegenomen als item [Check is the score matrix align with the number of items; it doesn't because source titles were used as items]		
1		Opzich werkt dit wel handig zo, als er steeds een winnaar in beeld komt. [vult ranks in t/m 11]	Automatische identificatie van winnaar: positieve ervring [Positive experience with the automatic identification of the winner]	Automatically showing the next winner in the tool is consedered healpful	-
2		Hebben we meerdere winnaars? [scrolt horizontaal] Nee. Oke, zit er in.	Controleert of de tool meerdere winnaars identificeert [Checks if the tool identifies multiple winners]	Checking for multiple winners in the tool	-
1		Hoe ga ik dat dan weer doen..? Dat staat denk ik niet hier. [opent scenario1] Hm, hier staat het niet. [sluit scenario 1] Oke. Sum-scores… Winnaar… [klikt tabblad ranking outcome aan] Ik had verwacht dat je hier dan wat in zou zien [klikt tabbladen labelslist, ranking outcome, verenigingen, ranking outcome aan] , maar dat gebeurt dus niet automatisch. [klikt tabblad voters aan] Voters… [klikt tabbladen labels, RRV aan]	Manier van handmatig ranken niet direct duidelijk [Method of manual ranking is not immediately clear]	It is generally not immediately clear where and how to rank in the tool without additional instructions	We will make the ranking more easy by automating the ranking process in the next iteration of the tool. This will ensure that it is no longer needed to manually assign rangs to winning items (except for ties) and that the user no longer needs to understand where and how to rank (except where to click a button for automated ranking and how to deal with ties).

2		[klikt op tabblad ranking outcome, klikt op tabblad RRV, klikt op tabblad ranking outcome] Kan je, [klikt op tabblad RRV] kan je nog één keer herhalen wat je vroeg? De, De… He?	Onduideljk waar er handmatig geranked moet worden [Unclarity about where th manual ranking takes place]		
2		Heb ik daar data voor gekregen? [scrollt verticaal] Even kijken... Dit zijn gewoon alleen maar gemiddelden. Hoe kan ik dit het gemakkelijkste doen? Ik ga eerst proberen te knippen en te plakken. [klikt op vergrendelde cellen] Even kijken, kan ik dat doen? Doet mijn Excel dat wel? Ik kan deze cel niet aantikken. [klikt op tabblad ranking outcome] Waarom werkt dat niet? Even kijken. [klikt op tabblad RRV] Misschien moet ik [klikt op vergrendelde cellen] dit [inaudible] doen. Wacht, kan ik die… [klikt op tabblad labelslist] hier in plakken? [selecteert en kopieert de modulenamen, klikt op tabblad ranking outcome, probeert de modulenamen te plakken in kolom B (gerankte modulenamen) te plakken maar krijgt een foutmelding] Nee, dat werkt niet.	Onduidelijk hoe er (handmatig) geranked moet worden [Unclarity about how manual ranking takes place]		
3		En dan heeft ‘ie al meteen al een winner aangewezen. [scrolt naar beneden] Ehm… Waar ik op zoek naar ben is ehm. D’r is nu steeds één winner [scrolt naar boven], maar je kan vast ook aangeven dat je een, een range wilt hebben van eh, weet ik veel, dat je de eerste vijf wilt meenemen [klikt op tabblad ‘voters’].	Manier van handmatig ranken is nog onduidelijk [Method of manual ranking is still unclear]		
4		Ja, ik zou dan hier [klikt op tabblad ‘ranking outcome’] spieken, maar dan doet ‘ie dus niks. [klikt op tabblad ‘RRV’] Ehm, maar, ja ik… ik.. (…) Ik ga gewoon even heel stom, ehh, tellen. [vult een 3 correct in] Maar nu schuift het op. Oh wacht, dat is dan drie. [vult 3 in op rij 2 bij een item die niet als winnaar werd geïdentificeerd] Ehh, dat is dan weer… [wijzigt de eerste 3 naar 4 in rij 2] Maar… Misschien. Vijf en zes [vult 5 en 6 niet correct in]. Ehm deze begint bij 68 en dan moet ik naar.. 59 [vult 7 in rij 2 niet bij winner in], 57 [vult 8 in rij 2 niet bij de winner in].. 56 [vult 9 in rij 2 niet bij de winner in], 54 [vult 10 in rij 2 niet bij de winner in]. Oh dan heb ik er twee met dezelfde dus dat wordt dan eh..	Manier van handmatig ranken is onduidelijk [Method of manual ranking is unclear]		
4		Maar ik snap, ik heb dit nu gebaseerd op dit he? [wijst met cursor naar rij 4 met de ongewogen som scores] Van totaal op die totaal scores, alleen ik ben, ik sta er niet helemaal achter en ik snap niet wat er nu gebeurd. Ja ik, dat ‘ie dat nu gaat	Manier van handmatig ranken is onduidelijk [Method of manual		

		delen denk ik ofzo, maar ik snap niet waarom dit [wijst naar de cel met 'winner' in rij 3] dan gaat verschuiven zeg maar. Dat die winnaar dan hier naartoe gaat [wijs met de cursor naar de kolom waar 'winner' in rij 3 staat].	ranking is unclear]		
2	[I]: Oke, ik zal ietsjes, ik zal ietsjes helpen in de.. in de richting. [P02]: Ja, [I]: Je ziet dat er een winnaar gekozen is, [P02]: Ja. [I]: En je kan boven die cel van de winnaar… [P02]: [dubbelklik op de cel boven de winnaar] [I]: Die cel inderdaad. Daar kun je zeggen van: oke, dit is dus de winnaar, deze heeft rank 1.	Ja. [vult 1 in de cel in] Uh ohw! Oke, maar nu is alles veranderd. (…) Hmmm, [klikt op tabblad ranking outcome] wat is er nu met ranking outcome gebeurd? Hey… Oke, maar hier is nu iets… Dat heeft 'ie daar dus daar zelf neer gezet. (…) [scrollt verticaal] Ik heb het gevoel dat ik nu zelf moet kijken welke dan.. nu, zeg maar, het hoogste gemiddelde heeft. Even kijken, 38 lijkt dan de hoogste. [scrollt horizontaal] Staat 'ie ook…? Nee hij staat niet op [inaudible]… Dat is deze dus nu. [selecteert de correcte cell voor de ranking, vult 2 in de cel boven één van de MULTI(2)] Wat is er nu gebeurd? [klikt op tabblad ranking outcome] Ranking outcome, ja dan staat die dus daar [wijst met cursor naar kolom patient score (kolom C), klikt op tabblad RRV]. Oke, uhm, ik ga gewoon proberen om.. [scrollt horizontaal] het aflopend gemiddelde tot 10 te ranken. Even kijken, 21,8.. 22 is het hoogste, 19, 15, 18.. Hier ook 22… 22,6. Oh wacht, er staat winnaar bij zelfs. [vult rank 3 in de correcte cel in] Wacht, misschien ben ik niet zo snugger. Ja, ohh kijk, dat is automatisch.[vult de ranks in de correcte cellen in] Vier, vijf, vijf modules geprioriteerd. Zes. [drukt op enter en de selectie verspringt naar beneden vanwege vergrendelde cellen] Waarom gaat 'ie naar onder toe? Even kijken, zeven. Acht. Negen. Tien.	Ontdekt na aanwijzing hoe en waar de ranking in de tool plaats vindt. [Discovers how and where manual ranking takes place in the tool]	Where and how to rank may become clear after a little instruction	Provide background infrmation to the tool's automatic ranking in an accompanying document
3		Eh, ja. Dan deze dus 4 [vult correct 4 in]. [scrolt horizontaal] Dit wordt dan 5 [vult 5 in]. Eh [vult 6 in], op deze manier kun je dus ook aangeven hoeveel modules je wil eh, prioriteren. Dus hoeveel modules je wilt [scrolt horizontaal] meenemen eigenlijk. Deze zijn, oh dit is eh, ja, tot hier.. [klikt eerste cel in rij 2 aan die geen te ranken module bevat] gaan de items. [scrolt horizontaal] Dat zijn 19 modules? (…) 7 [vult 7 in], [scrolt horizontaal] 8… [vult 8 in] Ho.. [vult 9 en 10 ook in]	Manier van ranken wordt iets helder door een richting in de opdracht [Method of ranking becomes somewhat clear through some direction in the assignment]		

3		Aha, oke [klikt op de juiste cel boven de winnaar in rij 2]. De rank als cijfer. Dus de winnaar wordt dan 1 [vult 1 in in de cel]. Ah, dat is het! Dit wordt dan 2 [vult 2 in in de juiste cel van een multi-winner, negeert MULTI(2) in rij 3]. Dan 3 [vult 3 in in de juiste cel]. Heel goed.	Manier van handmatig ranken wordt duidelijk door de opdracht maar negeert MULTI(2) [Method of manual ranking becomes clear through the assignment, but ignores MULTI(2)]	It is not immediately clear how to deal with items tied for the same rank in the tool without additional instructions	When using the automatic ranking, we will ensure that Excel prompts a message when there is a tie that needs manual assignment. Additional information on an option how to deal with ties will be provided in an accompanying document (heterogeneity analysis).
1		Oh wacht deze [selecteert de cel boven een MULTI(2)] is dan nu gelijk tweede. Krijgen ze dan allebei een 2? [vult 2 in boven de cellen van MULTI(2)]	Onduidelijkheid over gelijke ranks: welke rank in te vullen? [Lack of clarity about tied ranks: which rank to fill out?]		
4		Dan geef ik ze allemaal dezelfde s.. score [vult 10 in op rij 2 bij een item met dezelfde ongewogen totaalscore], want ze hebben een gedeelde plek. Geen idee.	Vult bij gelijke ranks hetzelfde nummer in [Fills out the same number for identical ranks]		
4		Oh! De ger.. oh. [klikt op tabblad 'ranking outcome'] Ik snap niet waarom er dan hier N/B komt? [wijst met cursor naar een item waar er een #N/B melding in Excel staat] Oh omdat ik 10 [klikt op tabblad RRV] had ik natuurlijk twee keer had gedaan omdat die allebij dezelfde score kregen. [klikt op tabblad 'ranking outcome']	Heeft bij gelijke rank twee keer dezelfde rank toegekend aan twee verschillende items en krijgt daarom foutmelding op ranking outcome tab [Assigned the same rank twice to different items which causes the error on the ranking outcome tab]		
3		Waar is die volgende winnaar nou gebleven? Hier. Ehm. Waar was ik? 6? [scrolt horizontaal] Nee 6 had ik gehad, 7 dus.	Handmatig ranken en zelf tel bijhouden van ranks kan foutgevoelig zijn [Manual ranking and keeping up with the numbered ranks	Scrolling horizontally to find winners and manually keeping track of the ranking order may prove	We will automate the ranking process in the next iteration of the tool.

			may be prone to errors]	to be difficult	
1		Max score is 10… Moet je dit dan nu zelf met de hand gaan doen? Dat lijkt mij nogal veel werk. Lijkt mij makkelijker als het automatisch kan gaan. Ehm… Maar wat ik zo kan zien staat nog nergens, automatisch. [wijst naar cellen met somscores van patiënten]	Wens voor automatisering van de ranking [Wish to automate the ranking method]	Wish for an automatic ranking system/method	We will automate the ranking process in the next iteration of the tool.
2		Maar wat ik [klikt op tabblad RRV] wel een beetje vreemd vind, is dat je dat je dus handmatig.. [selecteert de cel naast rank 1] dat moet doen. Ik snap niet zo goed waarom 'ie dan niet zelf automatisch die dingen van hoog naar laag kan sorteren.	Wens om ranking te automatiseren [Wish to automate the ranking]		
4		[klikt op tabblad 'voters'] Nee, hier moet ik wel even zeggen hoe veel mensen er stemmen. Dus dan… Van de <*wetenschappelijke vereniging 1*> is er één iemand [telt in dataset, vult 1 correct in in de tool], van de <*wetenschappelijke vereniging 2*> zijn er 2 [telt in dataset, vult 2 correct in in de tool], van de <*wetenschappelijke vereniging 3*> zijn er 4? [telt in dataset] Zou misschien handig zijn als je dat soort van automatisch uit kan vissen ofzo. [vult 1 correct in in de tool]	Wens om aantal voters te automatiseren [Wish to automate the count of the number of voters]	Wish for an automatic method to count the number of participating voter	It is possible to automate counting the number of voters. However, its purpose was to adjust the weighting method for individual weights. We have decided that we will use a different weighting method independent of the number of voters per organization. Therefore, the number of voters per organization does not have to be filled out in the next iteration of the tool.
1		Oke. [klikt op tabblad ranking outcome] Dus daar moet ik ook iets in kunnen doen. [klikt op tabblad RRV, selecteert gele cellen 'mark when ≤'] Oke, staat er ergens… De mean is vooral 5 als ik het even zo grofweg bekijk. [inaudible] .. 10.. Dit gaat over de variantie. De variantie is gemiddeld 8, 7, hmm 11, 8. [vult 8 in de cel 'mark when ≤' voor variantie] Als 'ie kleiner is dan 8, of als ie groter is dan.. [selecteert cel 'mark when ≥ voor variantie] Kan natuurlijk niet hetzelfde.. 8,1 [vult 8,1 in de vel 'mark when ≥' voor variantie]. Ach, [inaudible] 8,01 we [inaudible] risico [vult 8,01 in de cel 'mark when ≥' voor variantie en selecteert 'mark	Werking van de heterogeniteits-analyse meteen zelf ontdekt [Functioning of the heterogeneity analysis was discovered without instructions]	Use and functioning of the heterogeneity analysis is generally clear	-

		when ≥' voor std. deviatie]. Ja. Nu zie ik welke hoger dan 8 variantie hebben.. en lager dan 8.			
2		En volgensmij moet je hier [selecteert de cel 'mark when ≤' voor variantie] kunnen aangeven wat, eh. Dat je een kleurtje of een markering wil als er 'ie een bepaalde spreiding boven of onder een bepaald getal heeft. (…) Oke, ik zeg alles wat onder de 8 zit [vult 8 in de cell 'mark when ≤' voor variantie in], en wat uhm, ehh, wacht, nee dit is alles onder de 8. Ahh wacht [verwijdert 8 in de cell 'mark when ≤' voor variantie, selecteert 'mark when ≥' voor variantie]]. En dan hier doe ik [vult 8 in de cell 'mark when ≥' voor variantie in] alles boven de 8. [scrollt horizontaal] Hmm kijk, dat zijn er dus best wel veel. Dus je kan checken. (…) Ehm, dus je kan hier [verwijderd 8 en vult 10 in de cel 'mark when ≥' in] gewoon handmatig eh, zeggen: variatie meer dan 10. Dan zie je dat er nog best wel wat items zijn die een hoge … [scrollt horizontaal] variantie hebben. (…) Hmm, wat is dit? [wijst met cursos naar cellen 'mark when' voor kwartielgrootten] Oh dat gaat over de kwartielen. Oke, en wat is dit nog? [selecteert de cel 'mark when ≥' voor std deviatie] Doet dit ook nog iets? Dit is de, oh, de standaarddeviatie. [vult 1 in de cel 'mark when ≥' voor std deviatie]Oke.	Zelf relatief eenvoudig ontdekt hoe de heterogeniteitsanalyse werkt [Functioning of the heterogeneity analysis was discovered relatively easy without instructions]		
3		[scrolt naar boven] Ehh.. Ohja, hier [wijst met cursor naar de heterogeniteitsanalyse weergave van de items] geeft ie eh.. hier geeft 'ie daar informatie over. [klikt op gele cel 'mark when ≤' voor variance in rij 12] Oke, dus als de variantie groot is, eh en de standaard deviatie groot is, ehm.. dan kan ik dat hiermee [klikt op gele cel 'mark when ≤' voor standaard deviatie in rij 13], dan kan ik dat hiemee laten… even kijken hoor. Ehm. Even kijken welke nou een grote eh.. standaard deviatie hebben. [scrolt horizontaal even heen en terug][inaudible] Nou laten we bijvoorbeeld eens eh, als 'ie tussen de.. 3.. [vult 3 in gele cel 'mark when ≤' voor standaard deviatie] en 3.5 is. [inaudible] [vult 3,5 in rode cel 'mark when ≥' voor standaard deviatie]	Gebruikt uit zichzelf de heterogeniteits-analyse correct, zonder aanwijzingen [Correctly uses the heterogeneity analysis without instructions]		
4		Markeer wanneer het groter is dan… [klikt op cel 'mark when ≤', scrolt naar beneden, scrolt naar boven] dan… zes [vult 6 in bij 'mark when ≤' voor variantie] en kleiner dan 2 [vult 2 in bij 'mark when ≥' voor variantie]. Groter dan 6. Alles is eigenlijk groter	Gebruikt het principe van heterogeniteits-analyse correct, maar draait		

		dan 6. Nee maar waarom gaat deze dan, waarom wordt deze [wijs met cursor een rood gemarkeerde cel aan] dan rood. Deze is 4,75. En… kleiner dan… 1 [vult 1 in bij 'mark when ≥' voor variantie]. Huh? [verwijdert de 1 in de cel van 'mark when ≥' bij variantie] Markeer wanneer, nu wordt 'ie r… groter dan 2 [vult 2 in bij 'mark when ≤' voor variantie]. Maar dan zegt 'ie niks. Groter dan 6 [vult 6 in bij 'mark when ≤' voor variantie]. Wil gewoon groter dan [vult 8 in bij 'mark when ≤' voor variantie] acht, hier. He? Waarom wordt het dan 7,10 [wijst naar een geel gemarkeerde cel], dat is toch… kleiner dan 8, ohja. Kleiner dan… acht. [vult 8 in bij 'mark when ≤' voor standaard deviatie] Kleiner dan 8. Is alles. [verwijdert 8 in de cellen van 'mark when ≤' bij variantie en standaard deviatie] Kleiner dan 2 is bijna niks. [vult 3 in bij 'mark when ≤' voor variantie] Kleiner dan drie en of groter dan 8 [vult 8 in bij 'mark when ≥' voor variantie]. Kleiner dan 3 of groter dan 8? Wat was ook al weer de opdracht? Groter?	≥ en ≤ om in het gebruik. [Uses the principles of the heterogeneity analysis correctly, but confuses ≥ and ≤]		
3		Hoe groter de spreiding, ehm, ja, wat zegt dat? Dat mensen er heel verschillend over denken, in feite. Ehm. Ja, en wat ga je dan uiteindelijk met die informatie doen? Dat betekent in ieder geval dat er discussie over zal zijn… maar is dat dan een reden om 'm wel of niet te prioriteren?	Vraagt zich af wat het nut is van de heterogeniteits-analyse [Wonders what the use of the heterogeneity analysis is]	It may be unclear why one wants to use the heterogeneity analysis	We will provide additional information in an accompanying document about a possible role of the heterogeneity analysis (ranking with ties).
1		De lijst met ranking.. [klikt tabblad ranking outcome aan] Die had ik hier weer gezien. Dus volgens mij kan ik die heel makkelijk kopiëren [kopieert de ranking outcome] naar een.. Word-bestand. Word-bestand zei je he? [plakt de ranking in een Word bestand]	Correcte en eenvoudige output uit de tool [Correct and easy output from the tool]	Getting the output out of the tool is generally easy to perform	-
2		Oke. Ehh, [klikt op tabblad ranking outcome] ik ga naar het tabblad.. [selecteert de correcte cellen] van eh.. ranking outcome en ik ga dit gewoon selecteren en even [kopieert de tabel] control-c doen en dan kijk ik eventjes of ik dit in een Word-document kan plakken. Maar volgens mij kan je het niet zien dat ik dit nu open. Ja, ik heb het geplakt. Het staat gewoon in een word-document, dus ik kan het zo, eh..	Output: locatie voor output helder en daarna eenvoudige output uit de tool [Location for output of the tool was clear, and thereafter the output of the tool was easy]		
3		Oke. [klikt op tabblad 'ranking outcome'] Nou, dit is de ranking. Ehm. [selecteert de cellen met de header rank, name, patients-core en de bijbehorende gerankte items] Als ik dat gewoon copy-paste in Word… Dan krijg ik dus een lijstje met de volgorde waarin 'ie geprioriteerd is.	Output: eenvoudige manier om output uit de tool te krijgen [Easy output out of the tool]		

4		Nou, dan ga ik [selecteert en kopieert 13 items met rank / naam / patiënt score], dit zou ik gewoon kopiëren. Ehh… Ja gewoon plakken. [plakt de lijst in Word]	Output: eenvoudige manier van output uit de tool krijgen [Easy way of getting output out of the tool]		
4		Dat ik dit zeg maar [selecteert de eerste 5 items in de tool] even naar beneden… [verplaatst de geselecteerde items 1 rij naar beneden] Ik zal denk ik zo doen zeg maar [kopieert en plakt de richtlijntitel vanuit de dataset naar de tool in de lege rijd die door het verplaatsen is ontstaan], dan toch maar. (…) Ja, maar waarom krijg ik dan hier… [wijst met cursor naar een item in rij 2 met een #VERW! melding door Excel] Oh dat mag je natuurlijk niet zeggen, maar ik zou nu denken: eh, waarom krijg ik dan hier verwijderd?	Verplaatsen van cellen binnen de tool is niet te adviseren [Moving cells within the tool not advisable]	It is not advisable to move or cut/paste data within the tool to prevent errors	We will provide additional information in an accompanying document that will advise against moving or cutting/pasting cells within the tool.
4		Zo [verwijdert de richtlijn titel van richtlijn 2], zo [verwijdert de richtlijntitel van richtlijn 1, selecteert de module namen van richtlijn 1, knipt en plakt de geselecteerde modulenamen 1 rij naar boven, selecteert de module namen van richtlijn 2, knipt en plakt de geselecteerde module namen 2 rijen naar boven, verwijdert de naam van richtlijn 3, selecteert de modulenamen van richtlijn 3, knipt en plakt de geselecteerde modulenamen 3 rijen naar boven]. Zo, nu moet het wel kloppen denk ik [klikt op tabblad RRV]. Alleen nu zie ik hier 'verwijderd' [wijst naar #VERW! In rij 2 van een aantal items], maar dat is denk ik, eh, ahh.	Knippen/verplaatsen binnen de tool is niet te adviseren [Cut/paste within the tool is not advisable]		
4		Als je dan hier bijvoorbeeld invult [klikt op tabblad 'voters']: <*wetenschappelijke vereniging 1*> 1. Wat gaat er dan gebeuren? [klikt op tabblad RRV] Gebeurt er dan wat? [klikt op tabblad 'voters', vult 1 bij <*wetenschappelijke vereniging 1*> correct in, klikt op tabblad 'RRV'] Zou het even… [klikt op tabblad 'voters'] Is dat zinvol om te doen? (…) Ehhh… Nou dan moet ik bijvoorbeeld kijken waar dan meerdere <*wetenschappelijke vereniging 2*> mensen op gestemd hebben, dus dat is dan… Ohja, dat is natuurlijk bij alle modules. Ehm pff [klikt op tabblad 'voters'] zie ik wat veranderen? [klikt op tabblad RRV] Ehmm… 3.1, 3.2, 3.4 [klikt op tabblad 'voters'] Ja, [klikt op tabblad RRV] eh drie, ja volgens mij wel: 3.1 zei ik net 3.2, 3.4. [klikt op tabblad 'voters'] Als ik dan hier doe, [klikt op tabblad RRV] nu staat er: 3.3, 3.3, 3.5 [klikt op tabblad 'voters', klikt op tabblad RRV] / 3.3, 3.3, 3.5, nee heh? [klikt op tabblad 'voters'] Nee. [klikt op tabblad 'RRV']	Tool geeft geen feedback over wat er gebeurt met aantal voters en de invloed op de weging [The tool does not provide feedback about what happens with the number of voters and its influence on the weighting method]	The tool does not provide real-time feedback as to what happens with the number of voters and its influence on the weigthing method	We have decided that we will use a different weighting method independent of the number of voters per organization. Therefore, the number of voters per organization does not have to be filled out in the next iteration of the tool.

A3 Table 4 - Part 4: Using the tool in scenario 2

P	Additional instruction	Participant text	Label/observation	Interpretation	Actions to improve
1		Oke. Modulelijst… Krijg nou eerst alle modules weer… [kopieert en plakt module namen uit scenario 2 naar de correcte cellen in tabblad labels list] Dat zijn maar twee richtlijnen zo te zien. [kopieert en plakt overige module namen uit de lijst in de daarvoor bestemde cellen]	Input: zelfstandig itemnamen in tool gevoerd [intem names were independently entered into the tool]	Input of data into the tool is easy to perform	-
1		De naam… [kopieert alle namen uit scenario2] Het staat net andersom, na de vereniging. [plakt alle namen in de RRV tool in de correcte cellen] [selecteert en kopieert alle verenigingen uit scenario2] Kopiëren… [plakt alle verenigingen in de correcte cellen in de RRV tool] En dan kunnen we de scores overzetten… [kopieert alle scores in scenario2] En plakken. [plakt de scores in de correcte cellen]	Input: correcte input van deelnemers, organisaties, scores [Correct input of participants, organizations, scores]		
2		Ja, [plakt de modulenamen in de correcte cellen] even kijken, [scrollt verticaal] postoperatief beleid moet de laatste zijn die ik geselecteerd heb. Dan nog oesofagus [inaudible] carcinoom. Ehhm, [plakt modulenamen inde correcte cellen] die plak ik er ook in. [scrollt verticaal] Oke, even kijken. Uit scenario 2: ik heb alle modules, dan heb ik een database, ik heb een vereniging. Die ging ik eerst plakken. En.. [klikt op tabblad RRV] die ging ik op de RRV pagina doen, oeps. Even kijken bij [plakt de verenigingen in de correcte cellen] organisatie. En ik heb dan de namen van alle dokters. [plakt de deelnemersnamen in de correcte cellen] Die plak ik er ook in, en dan… moest ik nog iets plakken. Natuurlijk de scores. (…) [plakt alle scores in de correcte cellen]. Eh, alle scores staan er nu in.	Input: voert geheel zelfstandig direct alle data correct in de tool in [Immediately and independently fills out all input data correctly into the tool]		
3		klikt op tabblad 'labels list'] Oke [plakt item namen in de correcte cellen, scrolt naar beneden]. Dit zijn er ook echt, eh, wat meer. (…) [plakt een tweede set modules aansluitend in de lijst]. Even kijken, dit zijn alle modules [klikt op tabblad 'RRV']. Zo. En dan de… eh.. Ohja. De namen, en verenigingen. Ik begin met de namen. [inaudible] kopiëren. [plakt alle namen in de correcte cellen] Zo. Organisaties… Verenigingen.	Input: eenvoudige manier om input in de tool te krijgen [Easy to place input into the tool]		

		(…) plakt alle verenigingen in de juiste cellen] (…) Ehm, dan de scores. Ja, het zijn nog best wel wat modules. (…) [plakt de matrix met scores in de juiste cellen]			
4		Oke, en dan dit [wijst me cursor naar modulenamen in de database] moet je, dus dit is dan de item-lijst [selecteert en kopieert de modulenamen van de eerste richtlijn in de database]. (…) Hmm. [opent de tool, klikt op tabblad 'labels list, plakt modulenamen correct in de lijst, selecteert en kopieert de modulenamen van de tweede richtlijn in de database] Dit dan daar [plakt modulenamen correct in de lijst]. En dan deelname… [selecteert cellen met deelname aantallen in de dataset] Oh dit kan je dan, dit is nu dan wel nodig. Ehm.. die tool [opent de tool, klikt op tabblad 'voters'], met alle mensen die stemmen. <*wetenschappelijke vereniging 1*> 2 [vult 2 correct in], <*wetenschappelijke vereniging 2*> 1 [vult 1 correct in], <*wetenschappelijke vereniging 3*> 3 [vult 3 correct in], <*wetenschappelijke vereniging 3*>.. hee, ohja hier. 1 [vult 1 correct in]. <*wetenschappelijke vereniging 5*> 1 [vult 1 correct in], <*wetenschappelijke vereniging 6*> 1 [vult 1 correct in], <*wetenschappelijke vereniging 7*> 9 [vult 9 correct in], <*wetenschappelijke vereniging 8*> is 2 [vult 2 correct in], <*wetenschappelijke vereniging 9*> is 1 [vult 1 correct in], <*wetenschappelijke vereniging 10*> is 2 [vult 2 correct in], <*wetenschappelijke vereniging 11*> is 2 [vult 2 correct in], de <*wetenschappelijke vereniging 12*> is 1 [vult 1 correct in]. En dan moest ik [klikt op tabblad 'RRV'] dit nog even doen. Naam [selecteert en kopieert de kolom met namen uit de dataset, plakt deze correct in de tool]. Vereniging [selecteert en kopieert de kolom met verenigingen uit de dataset, plakt deze correct in de tool]. [selecteert en kopieert de matrix met scores uit de dataset, plakt deze correct in de tool] Scores.	Input: vult de input direct op je correcte plaats in de tool in [Fills out the input data corretly into the tool]		
1		[klikt tabbladen outcome ranking, RRV, outcome ranking aan] Even kijken, de ranking staat dan hier [selecteert en kopieert de gehele ranking in tabblad ranking outcome]… Pak gewoon weer even hetzelfde Word-bestand. Ik haal deze weer weg [verwijdert tabel van scenario1]. [plakt en formatteert de gerankte lijst in Word] Eh, als tabelletje vind ik fijner. Ja. [klikt tabblad RRV aan in de tool]	Output: eenvoudige output uit de tool [Easy output out of the tool]	Output of data out of the tool is easy to perform	-
1		[klikt tabblad ranking outcome aan] Even kijken, want zou eerst	Output: eenvoudige		

		gelijk hetzelfde als de andere lijst. [kopieert de hele ranking] Maar als ik 'm hier onder plak [plakt de ranking onder de tabel zonder penaltymethode in Word]… dan zou ik denk ik verschil moeten zien.	output uit de tool [Easy output out of the tool]
2		[klikt op tabblad ranking outcome] Oke. Even kijken, ik heb hier 15 eh.. [selecteert de correcte cellen voor de top15 om te kopieren] dingetjes geselecteerd. (…) [kopieert de correcte cellen] (…) Ik open even mijn word, eh-documentje. Ja, hij staat nu in Word.	Output: eenvoudige output uit de tool [Easy output out of the tool]
2		[klikt op tabblad ranking outcome, selecteert de correcte cellen] Oke, [kopieert de geselecteerde cellen] dan heb ik weer geen patiënten die, eh.. Plak ik in hetzelfde Word-bestandje.	Output: eenvoudige output uit de tool [Easy output out of the tool]
3		Ja. [klikt op tabblad 'Ranking outcome', selecteert en kopieert de gerankte modules (rank / name / patient score) zonder header] Dan copy-past ik dit even. Zo.. Ja..?	Output: eenvoudige output uit de tool [Easy output out of the tool]
3		Dit lijstje kopiëer ik dan naar mijn Word-bestand. [klikt op tabblad 'ranking outcome'] Ranking outcome. Zo [selecteert en kopieert alle ranks / names / patient scores onder header].	Output: eenvoudige output uit de tool {Easy output out of the tool]
4		[selecteert en kopieert de gerankte lijst met rank / name / patient score zonder header] Maar ik weet niet wat dat dan… Ehm… [plakt de gerankte lijst in Word] Eh, ik doe het gewoon even helemaal volgens het boekje, zo [herstelt de opmaak van de gerankte lijst in Word].	Output: eenvoudige output uit de tool [Easy output out of the tool]
4		En dan is het denk ik de bedoeling dat ik nu weer hier dit [selecteert de cellen met rank / name / patient score van de gerankte items], deze lijst kopieer/plak? (…) [kopieert de geselecteerde cellen] En dan daar onder, dan kan ik het goed met elkaar vergelijken straks. (…) [plakt de lijst met een andere opmaak in Word, verwijdert de lijst] Ho, [onverstaabaar, plakt de lijst met de oorspronkelijke opmaak in Word].	Output: eenvoudige output uit de tool [Easy output out of the tool]

4		[scrolt horizontaal, klikt op tabblad 'ranking outcome', selecteert en kopieert de cellen van items met rank/ name / patient score zonder header, plakt de lijst in word] Ehm… [verwijdert de lijst in word, plakt de lijst in word met de oorspronkelijke opmaak]] Zet ik 'm er naast, dan kan ik beter vergelijken.	Output: eenvoudige output uit de tool [Easy output out of the tool]		
1		Ja. [klikt op tabblad deelname in scenario2] En hier staat.. oh deelname. Mooi dan kan ik het mooi overnemen. <*wetenschappelijke vereniging 1*> heeft er 2 [vult aantal stemmers in], de <*wetenschappelijke vereniging 2*>… heeft er 1, <*wetenschappelijke vereniging 3*>…3. Dit zou misschien makkelijker moeten kunnen, maar..	Aantal voters invullen lijkt onhandig [Filling out the number of voters seems inconvenient]	Filling out the number of participating voters seems to be inconvenient en potentially prone to errors	We have decided that we will usr a different weighting method independent of the number of voters per organization. Therefore, the number of voters per organization does not have to be filled out in the next iteration of the tool.
2		En… [klikt op tabblad voters] Ik ga even naar de voters. Ah ja, en nu vind ik het wel onhandig dat ik dat handmatig bij de vereniging moet zoeken. Even kijken waar dat staat. <*wetenschappelijke vereniging*> zijn er 2, kan ik dit niet slimmer doen? <*wetenschappelijke vereniging*>.. [zoekt <*wetenschappelijke vereniging*> in de lijst] pfff. Hmmm. [vult 2 in de correcte cel in] Oh wacht, ik ga even mijn schermen naast elkaar zetten want dan kan ik het veel makkelijker lezen [wijzigt schermweergave in windows]. (…) Het staat op alfabetische volgorde dus dat is toch wel iets handiger, maar.. wel een beetje zoeken en ik denk ook ook zo'n beetje fout gevoelig.	Invullen van aantal voters lijkt onhandig en onverzichtelijk te zijn [Filling out the number of voters seems to be inconvenient]		
3		Ja. [inaudible] [klikt op tabblad 'voters'] Eh, ja dus <*wetenschappelijke vereniging 1*> twee dat klopt inderdaad. Dat staat daar ook. <*wetenschappelijke vereniging2*>.. eh, <*wetenschappelijke vereniging 2*> 1, <*wetenschappelijke vereniging3* > 3. <*wetenschappelijke vereniging 3*> 3…[vult 3 in de correcte cel in] <*wetenschappelijke vereniging 2*> 1.. [vult 1 correct in] Ehm.. <*wetenschappelijke vereniging 4*>, <*wetenschappelijke vereniging 5*> allemaal 1… [vult beide correct in] Ja. <*wetenschappelijke vereniging 6*> 5, <*wetenschappelijke vereniging 7*> 9 [vult beide correct in]. Ja zie je, de <*wetenschappelijke vereniging 7*> is inderdaad met veel en dat zie je ook wel echt terug in de uitkomst. En de <*wetenschappelijke vereniging 8*> 2	Handmatig invullen van aantal voters lijkt onhandig en foutgevoelig [Manually filling out the number of voters could be inconvenient and prone to errors]		

		[vult 1 in]. Hmm. Even kijken hoor, waar staat 'ie? [vult 1 in voor <*wetenschappelijke vereniging 9*>] Ehm, dit waren er 2 [vult 2 in voor <*wetenschappelijke vereniging 9*>]. Nee, wacht nu doe ik het verkeerd. 2 en 1. [vult 2 in voor <*wetenschappelijke vereniging 7*>, vult 1 in voor <*wetenschappelijke vereniging 9*>] Ehm. <*wetenschappelijke vereniging 10*> 2, <*wetenschappelijke vereniging 11*> 2… [vult 2 in voor <*wetenschappelijke vereniging 10*>. Hier, 2 [vult 2 in voor <*wetenschappelijke vereniging 11*>]. Het staat natuurlijk op alfabetische volgorde. En dan de <*wetenschappelijke vereniging 12*> nog 1, dat is de laatste. Hier. [vult 1 in voor <*wetenschappelijke vereniging 12*>] Oke. In totaal 30 [scrolt naar boven en naar beneden], dat klopt volgens mij ook nog [klikt op tabblad 'RRV']…. Met het aantal stemmers. Oke.			
1		Top, volgens mij staat het er zo in. [klikt tabblad RRV aan] Staat er dan hier [scrollt verticaal] ook iets bij? Niet dat ik zo kan zien. Oke. [klikt tabblad ranking outcome aan] Het staat er wel in nu. [klikt tabblad voters aan]	Tool geeft geen feedback over wat het doet met het aantal voters [Tool does not provide feedback regarding the number of voters]		
1		[selecteert de rij met ranks] Deze kan dus weer weg. [verwijdert de inhoud van de geselecteerde cellen] Is nu alles weg? [scrollt horizontaal, verwijderd laatste cellen met ranks] Nee. Ja hij is weer leeg, oke?	Verwijderen van de ranks in de tool moet handmatig gebeuren en lijkt onoverzichtelijk door horizontaal scrollen [Deleting the ranks is a manual task and seems inconvenient by scrolling horizontally]vv	Deleting ranks from the designated row in the tool seems inconvenient	A button to automatically clear all ranks will be developed in the next iteration of the tool, so that this won't have to be performed manually.
1		Dan vul ik hier 2 in [vult 2 in de correcte cel in]. En gebeurt er dan wat? [scrollt naar boven voor de somscores/ranking/winner] Oh, en dan moet ik misschien ook dit invullen [klikt op cel waar penalty aggression ingevuld kan worden], aggression.. Met 2.. null.. simple reweighted…	Tool geeft geen feedback over wat het doet bij invullen van penalty method en aggression [Tool does not provide feedback about how it uses the penalty aggression]	The tool does not provide real-time feedback when tool parameters change, since most formulas are calculated/ used in the	Because most formulas operate in the background of the tool it is dificult to provide real-time feedback when tool parameters are changed. We will provide additional

				background. There is a desire to see/know more about the penalty method and/or penalty aggression.	information on the penalty method and the penalty aggression in an accompanying document.
2		Wat is penalty aggression dan? Dat is… Only applicable to method 2, a penalty aggression of 0 equals a reweighted range voring method without weights. Wat dat is, echt geen idee? Maar ik kan dus volgens mij gewoon twee soorten penalties geven.	Geen achtergrond-informatie over penalty aggression [There is no additional information about the penalty aggression]		
2		[vult 1 in de correcte cel voor penalty method] Ik kan een eentje geven. Even kijken wat er dan gebeurt. [scrollt verticaal, scrollt horizontaal] Wacht even, nu weet ik niet meer welke module de winnaar was met de gewone penalty [vult 0 in de correcte cell voor penalty method]. Ik zet 'm even terug op 0. [scrollt virticaal, scrollt horizontaal] Dat was de herstadiëring. En als ik een ééntje geef [vult 1 in de correcte cel voor penalty method in] dan doe ik dus alleen op basis van statische groepsgrootte. [scrollt horizontaal] Dan krijg ik… een andere module… nee dezelfde. Shit ik weet het niet meer. Even kijken, [vult 2 in de correcte cel voor penalty methode] twee. [scrollt horizontaal] Ah, ik weet niet of ik nu niet scherp ben of dat ik nou niet goed zie. Wacht even, [vult 0 in de correcte cell voor penalty methode] nul.. (…) Maar dan lijkt er in de winnaar nu niks te veranderen met de penalty die ik geef. Waarschijnlijk dan andere dingen, maar dat heb ik dan niet gechecked. [klikt op tabblad ranking outcome] Kan ik dat bij mijn ranking zien?	Onvoldoende achtergrond over werking pentaly methode [Not enough information provided about the penalty method]		
3		En.. Dan zouden we dus kunnen spelen van: wat er gebeurt als je dan op 1, als je daar 1 van maakt of 2 van maakt, of.. nou, hoeveel, 23 van maakt.	Geen kader in de tool gegeven van de welke mate van aggression te kiezen [No background provided for which magnitude of aggression to use]		
3		Herstadiëring na neoadjuvant. Ehm.. Dat is nog steeds wel dezelfde winnaar.	Achtergrond-informatie over wegingsmethode is wenselijk (1e winnaar is altijd		

3			hetzelfde) [First winner is always the same. Background information about the weighting methods seem desirable]		
3		En neoadjuvante chemoradiatie was hier twee. En dat is hier … ook twee. Minimale invasieve oesofagusresectie is drie. Dat was ook al zo. Hé dit lijkt toch niet zo heel anders. Chemotherapie … chemoradiatie, endoscopische behandeling. De top 5 lijkt eigenlijk niet… ehm. Oh nee wacht. Type anastomose <*wetenschappelijke vereniging*> staat hier op 5. En hier [wijst met cursor naar #5 in de ranking] staat een andere op 5. Oke, dus de top-4 is eigenlijk hetzelfde nog steeds.	Geen kader in de tool gegeven van de welke mate van aggression te kiezen (door a=2 is er weinig verschil) [No background provided for which magnitude of aggression to use (a=2 causes little difference)]		
4		Maar is het dan de bedoeling dat, dat bedenk ik mu nu he [misklik waardoor het werkblad verschiet, scrolt horizontaal en verticaal terug], dat als je dit dan zo meteen gaat doen met je cluster, dat je dit dan zelf inschat wat dan de waardes zijn? Maar goed, dat mag je natuurlijk nu nog niet [onverstaanbaar] [verwijdert rank 1 uit rij2, werkblad nog steeds op 60% grotte] van oke, waar is dat dan op gebaseerd [verwijdert rank 2 uit rij 1] en hoe [vult 1 correct in op rij 2] moet ik dat dan straks kiezen? [vult 2 correct in op rij 2] En wat zegt dat [vult 3 correct in op rij 2], wat houdt het in?	Onduidelijkheid over welke waarde van de aggression ingevuld moet worden, tool geeft geen info [Unclear which value of aggression is needed, the tool does not provide information]		
1		Eén tot en met 15, oke. [vult rank 1 in] Het is een stuk lastiger zoeken als het er zo veel zijn…	Automatische identificatie winnaar: winnaar is lastig te vinden bij veel items [The automatic identification of the winner is hard to find when there are lots of items]	Ranking manually is more difficult when there are lots of items present in the tool. Participant sneed to scroll horizontally often to follow the changing winner after assigning a rank.	We will automate the ranking process in the next iteration of the tool. This way, the changing winner does not have to be followed in the work sheet and a rank is assigned by clicking on a button.

1		Oke, dit [selecteert de cel boven de winnaar] is volgens mij nog steeds dezelfde Winnaar [vult 1 in als rank boven de winnaar]. Een, twee, drie, [vult ranking verder in] Ik wilde het veel te snel doen en dan is het frustrerend je niet gelijk die… Vijf, zes, zeven, acht, negen, volgens mij is het nu wel anders als eerst. Elf, twaalf, kijken.. dertien, veertien. En vijftien waar ben je? Daar. Yes.	Handmatig ranken lijkt onoverzichtelijk te zijn (horizontaal scrollen bij veel items en zelf ranks tellen) [Ranking mannually seems inconvenient (scrolling horizontally when there are lots of items)]
2		Zoomen, [zoomt uit in het Excelbestand naar 40%] kan ik misschien iets sneller zien waar de winnaar zit.	Handmatig ranken bij veel modules is mogelijk onoverzichtelijk [Manual ranking with lots of items is possibly inconvenient]
3		Het is nog best wel een klusje eigenlijk [vult 6 correct in]. Zeker als het een hele grote.. eh [scrolt horizontaal, vult 7 correct in]], een heel groot cluster is. Nouja hoewel [vult 8 correct in], aan de andere kant, ehm, hangt het natuurlijk ook af van hoeveel modules je kan prioriteren [scrolt horizontaal]. En dan valt het wel mee hoe groot dit klusje is. 9 [vult 9 correct in, scrolt horizontaal]	Handmatig ranken kan misschien een flinke klus zijn bij veel items [Ranking manually might be a big chore with lots of items]
3		[scrolt horizontaal, vult 2 correct in] Ho, ik klik iets te ver door [klikt verkeer op de scoll bar]. Ben de tel kwijt [lacht]. Drie… Volgens mij was het drie. Wat gebeurde er nou? Drie [vult 3 correct in, scrolt horizontaal], vier [vult 4 correct in], vijf [vult 5 correct in, scrolt horizontaal, vult 6 zin, scrolt horizontaal]… Zeven [vult 7 correct in, scrolt horizontaal], acht [vult 8 correct in], negen [vult 9 correct in], tien [vult 10 correct in, scrolt horizontaal], elf [vult 11 correct in], twaalf [vult 12 correct in, scrolt horizontaal, vult 13 correct in]… veertien [vult 14 correct in], en de laatste. Vijftien [vult 15 correct in].	Veel scrollen om de verspringende winnaar te volgen [Lots of horizontal scrolling to follow the changing winner]
4		[scrolt horizontaal] Oh, dat gaat dan… Oh! [vult 2 correct in op rij 2, scrolt horizontaal] Oh, dan moet je me maar een keer uitleggen hoe dat dan gaat, maar goed. Eh! [lacht, vult 3 correct in op rij 2] (…)	Veel horizontaal scrollen om verspringende winnaar te volgen

		Vier [vult 4 correct in op rij 2], dat gaat iets sneller dan toen net [vult 5 correct in op rij 2]. Zes [vult 6 correct in op rij 2, scrolt horizontaal]. 7 [vult 7 correct in op rij 2], 8 [vult 8 correct in op rij 2], 9 [vult 9 correct in op rij 2], 10 [vult 10 correct in op rij 2, scrolt horizontaal], ehmm, 11 [vult 11 correct in op rij 2, scrolt horizontaal]. He, 12 [vult 12 correct in op rij 2, scrolt horizontaal]. 13 [vult 13 correct in op rij 2, scrolt horizontaal]. 14 [vult 14 correct in op rij 2] vijftien [vult 15 correct in op rij 2, vult 16 correct in op rij 2].	[Lots of horizontal scrolling to follow the changing winner]		
4		Ehm nou ik ehm, jaa. [verkleint het werkblad naar 60%] Ik denk nu gewoon vooral zo snel mogelijk invullen [scrolt horizontaal, vult 9 correct in op rij 2, vult 10 correct in op rij 2]. Heel veel meer dan dat denk ik niet, denk ik [vult 11 correct in op rij 2].	Verkleint het werkblad om de verspringende winnaar te kunnen volgend voor de handmatige ranking [Zooms out in the work sheet to follow the changeing winner fort he manual ranking]		
2		Ik heb hier weer de gewogen gemiddelden en ik heb een winnaar [scrollt horizontaal]. Ehm, ik ben alleen even vergeten: moest ik dit nu ook gaan ranken als nummer 1? [vult 1 in de correcte cel voor de ranking in (boven 'WINNER')] (…) Oke. [vult de ranks in de correcte cellen in] Twee… [scrollt horizontaal]Waar zit nummer 3. Whoops, 3. Vier, Vijf,	Voert ranking zelfstandig en correct uit [Carries out het ranking independently and correct]	Carrying out the ranking is generaly easy with the prior experience of getting to know the tool/ scoring / scenario 1.	We will automate the ranking process in the next iteration of the tool and provide guidance in an accompanying document.
2		Oke. Ehm, even kijken. [vult de ranks in de correcte cellen in] Een eentje. (…) Even.. Drie, volgens mij is dit inderdaad een hele andere nummer drie. Want dat was volgens mij net een van de laatste. Vier. Vijf. Zes. Zeven. Acht. Negen. Tien. Elf. Twaalf. Dertien. Viertien.	Vult de ranking uit zichzelf correct in [Independently completes the ranking correctly]		
3		scrolt horizontaal Ja, dat is 1 [vult 1 correct in]. [scrolt horizontaal] 2 [vult 2 correct in]. Hmm, [scrolt horizontaal] 3 [vult 3 correct in]. 4, 5 [vult 4 en 5 correct in]. Het is nog best wel een klusje eigenlijk [vult 6 correct in]. Zeker als het een hele grote.. eh [scrolt horizontaal, vult 7 correct in]], een heel groot cluster is. Nouja hoewel [vult 8 correct in], aan de andere kant, ehm,	Voert de ranking correct uit [Performs the ranking correctly]		

		hangt het natuurlijk ook af van hoeveel modules je kan prioriteren [scrolt horizontaal]. En dan valt het wel mee hoe groot dit klusje is. 9 [vult 9 correct in, scrolt horizontaal] Tien. [vult 10 correct in, vult 11 correct in, scrolt horizontaal]. 12 [vult 12 correct in, scrolt horizontaal, vult 13 in, scrolt horizontaal, vult 14 correct in]. 15… [scrolt horizontaal] hier, 15 [vult 15 correct in].			
3		Dan is dit 1 [vult 1 als rank correct in in de cel boven de winnaar]. Dan willen we natuurlijk het verschil zien tussen deze top-15 en de vorige top-15. Ehm… [scrolt horizontaal, vult 2 correct in, scrolt horizontaal, vult 3 correct in, vult 4 correct in, scrolt horizontaal] Vijf [vult 5 correct in].	Voert de ranking correct uit [Performs the ranking correctly]		
3		[scrolt horizontaal, vult 2 correct in] Ho, ik klik iets te ver door [klikt verkeer op de scoll bar]. Ben de tel kwijt [lacht]. Drie… Volgens mij was het drie. Wat gebeurde er nou? Drie [vult 3 correct in, scrolt horizontaal], vier [vult 4 correct in], vijf [vult 5 correct in, scrolt horizontaal, vult 6 zin, scrolt horizontaal]… Zeven [vult 7 correct in, scrolt horizontaal], acht [vult 8 correct in], negen [vult 9 correct in], tien [vult 10 correct in, scrolt horizontaal], elf [vult 11 correct in], twaalf [vult 12 correct in, scrolt horizontaal, vult 13 correct in]… veertien [vult 14 correct in], en de laatste. Vijftien [vult 15 correct in].	Voert de ranking correct uit [Carries out the ranking correctly]		
4		[vult 4 correct in op rij 2, vult 5 correct in op rij 2, vult 6 correct in op rij 2, vult 7 correct in op rij 2, vult 8 correct in op rij 2, vult 9 correct in op rij 2, vult 10 correct in op rij 2, vult 11 correct in op rij 2, vult 12 correct in op rij 2, vult 13 correct in op rij 2, vult 14 correct in op rij 2] Oke [vult 15 correct in op rij 2], 15, he?	Correcte uitvoering van de handmatige ranking [Carries out the manual ranking correctly]		
4	[I]: Je mag eerst een top-15 maken, door de rank als cijfer boven de winnaar te zetten. Dus je ziet, ehm, en dat is even.. Nu ga ik ook eh, even wat anders doen dan in de vorige scenario 1. Nu zieje, zeg maar, 'winner' staan. [P04]: [scrolt horizontaal] Ja? [I]: Daar oven mag	[scrolt horizontaal] Oh, dat gaat dan… Oh! [vult 2 correct in op rij 2, scrolt horizontaal] Oh, dan moet je me maar een keer uitleggen hoe dat dan gaat, maar goed. Eh! [lacht, vult 3 correct in op rij 2] (…) Vier [vult 4 correct in op rij 2], dat gaat iets sneller dan toen net [vult 5 correct in op rij 2]. Zes [vult 6 correct in op rij 2, scrolt horizontaal]. 7 [vult 7 correct in op rij 2], 8 [vult 8 correct in op rij 2], 9 [vult 9 correct in op rij 2], 10 [vult 10 correct in op rij 2, scrolt horizontaal], ehmm, 11 [vult 11 correct in op rij 2, scrolt horizontaal]. He, 12 [vult 12 correct in op rij 2, scrolt horizontaal]. 13 [vult 13 correct in op rij 2, scrolt horizontaal]. 14 [vult 14 correct in op rij 2] vijftien [vult 15 correct in op rij 2, vult 16 correct in op rij 2].	Manier van ranken lukt na aanwijzing [Method of ranking is clear after additional instruction]	Some additional information may still need to be provided about where and how to rank in the tool, whereafter ranking can be carried out correctly.	We will automate the ranking process in the next iteration of the tool and provide guidance in an accompanying document.

	je de rank telkens zetten. [P04]: [klikt op de cel in rij 2 boven 'winner'] [I]: Dus dit is 1. [P04]: Ohhh! [vult 1 correct in op rij 2] [I]: En dan gaat 'ie verspringen en dan de volgende winnaar is 2. Die heeft de tweede rank.				
4		[vult 1 correct in op rij 2, scrolt horizontaal, vult 2 correct in op rij 2, vult 3 correct in op rij 2, scrolt horizontaal, vult 4 correct in op rij 2, scrolt horizontaal, vult 5 correct in op rij 2, vult 6 correct in op rij 2, vult 7 correct in op rij 2, scrolt horizontaal, vult 8 correct in op rij 2]	Ranken lukt meteen na een eerdere aanwijzing tijdens scenario 2 [Carrying out the ranking correctly after prior intructions in scenario 2]		
2		[drukt op enter zodat de selectie verschiet naar de eerstvolgende onvergrendelde cel] Hey.. Waarom verschiet 'ie nou de hele tijd naar beneden? Huu.	Verspringen van selectie naar onbeveiligde cellen (door enter) [Pressing enter causes the selected cel lto jump tot he next unprotected cell]	Securing cells in het worksheet allow only some cells to be selected. Pressing enter will switch to the next available cell.	We keep the option to secure cells to guide the user through the available cells. We will add additional information to an accomanying document how to disable secured cells.
2		Mijn gedachten hierbij is dat het wel, best wel veel handmatig veel werk is wat je volgens mij ook automatisch door Excel zou moeten kunnen laten doen [scrollt horizontaal]. Als het toch al zo mooi automatisch kan.	Wens om de ranking te automatiseren [Wish to automate the ranking]	Wish to automate the method of ranking because manually ranking is considered to be a lot of work.	We will automate the ranking process in the next iteration of the tool.
4		En kan ik dan niet dit [selecteert de cel voor penalty method B29 in de tool parameters] dan op 2 aanpassen, en dat 'ie dan.. (…) [scrolt horizontaal] Maar als ik dan, oh dan.. Dan zijn die, dan is dit [selecteert een aantal cellen in rij 2 met de ranks van items]	Verwachting dat de eerder ingevulde ranking direct wijzigt als er een andere	Some might expect that the tool automatically adjusts all	It is dificult to provide real-time feedback when tool parameters are

		natuurlijk niet meer kloppend, want dit weegt waarschijnlijk mee in hoe 'ie dan de winnaar beoordeelt. Dus dan moet je dat eerst leeg maken [selecteert een reeks cellen met ranks in rij 2 en verwijdert de inhoud], denk ik.	penalty methode wordt toegepast [Expectation that the filled out ranking immediately changes when selecting a different penalty method]	prior filled out rankings when tool parameters change. However, this is currently not the case.	changed. We will develop a butten to clear all rankings at once and provide additional guidance about changing tool parameters when a ranking was performed.

A3 Table 5 - Part 5: Semi-structured interview

P	Participant text	Label/observation	Interpretation	Actions to improve
1	Maar op zich is het… goed dat als je daar dus een penalty voor toepast dat de ranking ook wel veranderd. Dat er dus wel rekening mee wordt gehouden dat mensen ook echt volgens hun eigen belangen stemmen. Zeker als er.. Volgens mij was er 1 vereniging waar er 5 mensen aan meededen, ja, die modules worden dan automatisch hoog gezet. Ik denk dat dit een mooie manier is om mee te nemen.	Ziet nut van optie voor penalty, disproportionele representatie [Sees use for a panalty: disproportionate representation]	There are some considerations about whether and when to use (dis)proprotional representation methods	We replace the current adjusted weighting methods (based on the number of voters) with another adjusted weighting method. There are pros and cons to both proportional en disproportional representation, depending on the context. We aim to keep both types of weighting methods in the tool.
1	Ik denk dat dat wel eerlijk is, want volgens mij was van de <*wetenschappelijke vereniging*> waren er 5 die gestemd hadden als ik het mij goed.. ik mij niet vergis. Dus ik denk dat het beter is dat je dan dus ook andere verenigingen modules toebedeeld.	Proportionele representatie lijkt niet altijd per se wenselijk [Proportional representation does not always seem tob e desirable]		
2	Nou ik, ja, ik, i-ik vraag mij eigenlijk een beetje af in hoeverre dat zeg maar wenselijk is. Ehm…, wat ik mij wel kan voorstellen, dat je inderdaad situaties hebt waarin je… ehm, met je belangrijke partners, eh.. in dit geval is de <*wetenschappelijke vereniging*> waarschijnlijk de hoofd WV en is het met je partners waarschijnlijk belangrijk dat je samen met je partners, zeg maar, samen gaat prioriteren maar tegelijkertijd… eh, als ik denk aan mijn eigen cluster benigne gynaecologie, dan zijn zeg maar… ja dan kan ik mij ook voorstellen dat het voornamelijk alle gynaecologische modules zijn die geprioriteerd worden, omdat dat, ja eh.. het doel van het cluster denk ik ook dient. Dus ik weet niet zo goed of dit… eh, per se de oplossing is om tot een betere prioritering te komen.	Afweging proportional vs disproportional representation [Consideration for proportional vs disproportional representation]		

2	Ehm, het heeft ook het voordeel dat je, bijvoorbeeld, voor een kleine partij als de <*wetenschappelijke vereniging*>, die vaak maar met 1 persoon in zo'n cluster zal zitten, dat je wel je stem, eh, kan inbrengen. Zeker als je in zo'n groot chirurgisch cluster zit. Ehm.. Maar ja… Nah ik weet ook niet zo goed wat je wel als prioriteringsmethode… Misschien moet je.. Ja, afspraken maken afhankelijk van het gewogen… He, hoeveelheid…	Afweging proportional vs disproportional representation [Consideration for proportional vs disproportional representation]		
2	Wat ik gewoon nu nog niet zo goed zie in dit voorbeeld is, het lijkt nu zeg maar alsof er soort van heel mooi divers palet, zeg maar, in het tweede scenario eh, ge..geranked is. Maar als dat helemaal niet reflecteert dat, zeg maar, 95% van het cluster puur <*wetenschappelijke vereniging*> modules zijn, dan ..voelt het meer alsof je een soort van dansje doet om alle kleine partijen tegemoet te komen dan dat je.. recht doet aan de inhoud van je cluster.	Afweging proportional vs disproportional representation [Consideration for proportional vs disproportional representation]		
3	Ehm, en eh, de groepsgrootte van een aantal afgevaardigden vanuit de WV zou er niet al te veel invloed op moeten hebben. En het lijkt erop dat dat als je dat beter spreidt, dat als je wel een penalty toevoegt, dat dan in ieder geval meer WVen hun zin krijgen. En tegelijkertijd zijn er natuurlijk veel argumenten te bedenken waarom je module prioriteert ja of nee, behalve alleen je WV belang.	Afweging proportional vs disproportional representation [Consideration for proportional vs disproportional representation]		
4	..dat is natuurlijk meer bij die eerste. En dus dan denk ik dat dat dus wel beter verdeeld is hier, want dan is dit, hier zie, bij die tweede, bij die zeg maar die eerste van die ranking, dus de tweede, daar zijn er maar twee modules van de <*wetenschappelijke vereniging*> geprioriteerd. En bij die laatste ook twee, drie, iets meer, vier. Ja, dus dat is wel beter verdeeld dan bij de, zeg maar, de echt, wat we helemaal als eerste deden. (…) In verhouding? Ja dan vind ik dat wel terecht, denk ik. Ehm, omdat anders wel is ehm, de meeste stemmen gelden. Aan de andere kant… Ehh, Ja, als er meer mensen van dezelfde vereniging in een werkgroep of, zeg maar, in een clusterwerkgroep zitten, dan wil je denk ik wel voorkomen dat er, dat.. dat is dan de meeste st… dat.. dat dat zeg maar.. dat dat zeg maar, dat uiteindelijk alleen maar de modules, dat de modules van die partij alleen, uiteindelijk alleen geprioriteerd worden. Aan de andere kant wil je dan ook weer niet helemaal dat het, zeg maar, is dat het… stel het is wel een cluster wat wel heel chirurgisch is, net zoals bijvoorbeeld traumatologie bij mij, dat het dan, dat dan, dat het dan zo streng gerekend wordt dat er dan uiteindelijk nog maar heel weinig komen vanuit de <*wetenschappelijke vereniging*> en dan heel veel andere modules. Want, dat het dan ook wel een heel chirurgisch onderwerp is, zeg maar. Dus dat het dan daarom veel chirurgen zijn	Afweging proportional vs disproportional representation [Consideration for proportional vs disproportional representation]		

1	Het is wel een beetje een black-box nu voor mij, van nou.. we vullen 2 en 4 in en er komt wat anders uit.	Tool geeft geen feedback over de berekening van het individuele gewicht [Tool does not provide feedback in the calculation of individual weights]	The tool does not provide feedback when changing tool parameters or when calculating individual weights	Since most formulas run in the background, it is dificult to provide real-time feedback. We will provide additional information in an accompanying document
1	Ja ik zou zelf wat meer achtergrond willen in hoe het wordt berekend, maar dat is misschien als onderzoeker dat je dat meer wilt weten van: hoe dan? Ik weet niet, staat het ergens in de tool ofniet?	Tool geeft geen feedback over de berekening van het individuele gewicht [Tool does not provide feedback in the calculation of individual weights]		
1	Ja, want jij zei bij die penalty aggression: vul daar maar 4 in. Maar ik heb nergens kunnen zien waar tussen je kon kiezen ofzo. En dat is misschien fijn als daar net nog als bij dat penalty method iets achter staat. Wanneer je wat moet invullen.	Tool geeft geen feedback over het gebruik van de penalty methode/ aggression [The tool does not provide feedback about the use of the penalty method and aggression]		
4	Ja dus dan, ja maar de verschillende was dat dan nummer 2 en 3? Ja ik, vind het dan heel lastig om te zien, zeg maar wat dan nummer drie doet. En ik heb voor mijzelf dan, denk ik, nu niet echt, dat het best wel snel ging voor mij dat ik dacht van: het beeld van hoe dan het verschil tussen 2 en 3 tot stand is gekomen om dan een goede mening te formuleren wat ik dan van 2 vond.	Tool geeft niet direct feedback over penalty methode [Tool does not provide feedback about the penalty method]		
1	Je vult alles in en er komt vanzelf een winnaar uit. Ik dacht eerst van: oh dat moet je nu zelf gaan berekenen, maar de winnaar had ik nog even niet gespot. Maar op zich, dat groene vlak valt heel goed op waardoor je heel snel de ranking kan toepassen. Dat werkt wel heel prettig, ja.	Automatisering in de tool wordt op prijs gesteld [Automation in the tool is being appreciated]	Automation within the tool is being appreciated	-

1	Die ranking, ja die wordt ook automatisch berekend als je die score hebt ingevuld, dus dat werk heel prettig.	Automatisering van detectie winnaar wordt op prijs gesteld [automatic detection of the winner is being valued]		
4	En ik vond het op zich wel goed dat je dan, kijk normaal heb je natuurlijk nog instructie bij hoe zo'n tool werkt, maar dat je dan die winner, dat je dan zo die, die eh, ranking zelf in kan vullen, maar dat je daar zelf verder niks voor hoeft te berekenen, dat 'ie dat dan automatisch volgt. Ja…	Automatische identificatie van de winnaar is handig [Automatic identification of the winner is helpful]		
1	Ja het werkt wel heel prettig. Ja. Het zou mooi zijn als je scherm, als je dit aan het invullen bent, want zeker als er straks 100 zijn ben je best wel aan het scrollen, dat 'ie automatisch naar het stukje gaat waar dan winnaar staat. Dat.. Dat maakt 'm nog makkelijker om te gebruiken, maar ik weet niet of dat mogelijk is in Excel. Dat soort dingen.	Wens om de ranking te automatiseren, anders veel scrollen [Wish to automate the ranking method, otherwise it is a lot of scrolling]	There seems to be a wish to automate the ranking method in order to avoid acrolling through the worksheet]	We will automate the ranking process in the next iteration of the tool.
2	Dat handmatig de ranking invullen, dat zou ik automatisch door Excel, zeg maar, laten doen. Dat je dat naar je tabblad laat, eh… invullen zeg maar. Dat lijkt mij eigenlijk wel makkelijk, want dat zit je niet de hele tijd te tutten, en dan kan je niet een fout maken dat je net een getal mist, ofzo.	Wens om ranking te automatiseren [Wish to automate the ranking method]		
2	Ja, maar dan zou je wel, eh.. een filter er in kunnen zetten en gewoon aanklikken: sorteer van hoog naar laag, bijvoorbeeld. En dan doet 'ie het automatisch, denk ik, voor je dat je dan.. ehm, je hoogst scorende module bovenaan krijgt.	Wens om ranking te automatiseren [Wish to automate the ranking method]		
3	Ja, dat zou misschien ook een goede toevoeging zijn. Dan moet je dus wel goed kunnen aangeven tot waar je wilt ranken. Eh, ja… tegelijkertijd houd je misschien iets meer feeling er mee als je het handmatig doet? Ja, ik weet niet of dat heel veel zou toevoegen eigenlijk. Je ziet natuurlijk ook de volgorde vanzelf.	Twijfel of automatisch ranken zorg voor verlies van gevoel voor data [Doubt whether automated ranking resukts is loss of felling with the data]	There might be some worries that automated ranking may cause a loss of feeling with the data	We will automate the ranking process in the next iteration of the tool. This will be done semi-automatically (by pressing a button to assign the next rank) in order to be able to trace the ranking process.
2	Maar als je natuurlijk heel veel modules hebt, zoals ik bij benigne gynaecologie hebben wij 186 modules, uit mijn hoofd. Ja, dat werkt natuurlijk niet als je die ranking [lacht] moet gaan zoeken, zeg maar, eh, eh, van waar nou de winnaar zit. Dus dan is het misschien toch	Overzicht om handmatig te ranken bij veel modules ontbreekt	Overview is lacking when manually assigning ranks in the tool when there are	We will automate the ranking process in the next iteration of the tool. This will reduce the need for horizontal scrolling

	handiger om die modules onder elkaar te listen. Om het overzicht te krijgen, zit ik mij zo gauw te bedenken.	[There is a lack of overview when assigning ranks manually if there are lots of items]	lots of items	in the worksheet (except for ties).
2	Nee, nouja ik denk, nee. Behalve dat ik dus denk dat de uitdaging vooral ligt van hoe je dat.. hoe je dit nu overzichtelijk houdt, als je dus heel veel modules hebt en dat je dan handmatig die ranking moet doen. Ik dat dat gewoon niet… ik denk dat dat niet werkt.	Handmatig ranken lijkt onoverzichtelijk indien er veel items in de tool staan [Ranking manually is inconvenient when there are lots of items]		
3	Ehm, het maken van die ranking, dat is nog wel eventjes een soort uitzoekwerkje.	Manier van ranken is uitzoeken [Manner of ranking is something to figure out]		
4	Want ik heb straks eh, volgens mij zijn het er 2… ik noem 200 ofzo, maar dat zijn natuurlijk enorm veel kolommen. Ehm, en ook als je dan zeg maar zelf de, die, die ranking moet doen. Dan moet je natuurlijk wel de hele tijd scrollen van links naar rechts.	Veel horizontaal scrollen bij handmatige ranking [Lots of horizontal scrolling when ranking manually]		
1	Nou, die verenigingen hebben we volgens mij niks mee gedaan. Alleen die voters zat ik een beetje van, want dan ben je aan het zoeken. Ehm, maar ik weet niet of dat makkelijker kan op een andere manier. Want nu zijn er heel veel partijen die je eigenlijk niet nodig hebt. Misschien zou je nog.. Want hier hebben we natuurlijk… Je ziet natuurlijk mijn scherm niet meer. Eh, het eerste tabblad, dat RRV, daar heb je nu organisaties ingevuld. Dat zou ergens makkelijk zijn als 'ie dat automatisch dan ook bij het aantal voters optelt. Dan hoef je het niet meer zelf in te vullen. (…) Maar dat maakt in ieder geval het tabblad voters nog weer gebruiksvriendelijker	Automatisering van invullen aantal voters zou wenselijk zijn [Automation of counting the number of participating voters would be appreciated]	Thereis a wish to automate the counting of the number of participating voters	We have decided that we will use a different weighting method independent of the number of voters per organization. Therefore, the number of voters per organization does not have to be filled out in the next iteration of the tool.
2	Maar ik weet niet of dat onder dit kopje valt: maar bijvoorbeeld bij het kopiëren van hoeveel mensen er in een vereniging zitten uit jouw cluster, zou je het misschien makkelijker kunnen maken door ofwel, ehm, de lijst waaruit je kunt kopiëren dat dat eigenlijk altijd een standaard lijstje is die je er gewoon overheen kunt plakken. Ehm, of iets anders dat je niet… nou goed, dat je zo min mogelijk typfouten, zeg maar, kan	Wens om het invullen van het aantal voters te automatiseren [Wish to automate the counting of the participating voters]		

	maken bij het overtypen. Dat is misschien nog handig.			
3	Eh, ja ik vroeg mij af of het niet in te bouwen is dat het aantal voters per vereniging, dat dat ook automatisch berekend wordt vanuit het RRV-tabblad. Want je geeft… Ja misschien is dat wel lastiger te programmeren, maar je geeft aan dat er een naam achter een WV komt, dat zou je dan toch wel moeten kunnen optellen. Dat zou misschien het gebruiksgemak nog wel vergroten, in plaats van dat je dat nog handmatig invoegt.	Wens om het invullen van het aantal voters te automatiseren [Wish to automate the count of participating voters]		
1	Zeker als er dan straks, eh, als je dus die penalty toepast en dat het dan in een keer anders wordt. Ik kan mij voorstellen dat de <*wetenschappelijke vereniging*> dan opeens denkt: we hadden eerst negen modules en nu nog maar 5. Dat je dan kan uitleggen hoe dat gedaan is. Ik weet niet of mensen gelijk genoegen nemen met: ik heb een penalty toegepast en daarom zijn er nu minder modules.	Achtergrond-informatie over penalty nodig om keuzes en uitkomsten te onderbouwen [Background information about penalty methods is needed to provide an argument for choices and outcomes]	There is a lack of background information about how the tool works and there is a lack of guidance which method and aggression to use.	We will provide additional information in an accompanying document about the workings of the tool, including the penalty methods and penalty aggression.
1	In principe wel, alleen ik zou dan de achtergrond van die penalty methode willen weten.	Achtergrond-informatie over penalty methode wenselijk [Background information about the penalty method seems desirable]		
2	Ja, ehm… Ik moet eerlijk zeggen dat ik het eigenlijk nog niet zo goed volg. Ik zit… Ik heb het Word-bestand er nog even bij geopend. (…) Maar het gaat mij een beetje te snel om te zien waar precies de verschillen in zitten, maar ik zie inderdaad wel er v.. dat er gewoon andere modules eh,tussen de twee dingen zitten, maar ik snap niet , ik snap nog niet wat ik gedaan heb met die penalty.	Werking penalty methode nog onduidelijk [It is unclear how the penalty method works]		
2	Ik denk.. Het is een mooie tool. Ehm, ik snap nog niet helemaal hoe die werkt.	Werking/ achtergrond tool onduidelijk [Mechanism / background of the tool is unclear]		
2	Misschien het laatste wat me als laatste nog te binnen schiet is dat op die RRV, eh, dat eerste tabblad, daar staat natuurlijk super veel data, ehm… waar je misschien wel helemaal niks mee doet en dan… weet	Achtergrond-informatie is wenselijk om de tool te begrijpen		

	ik niet zo goed of dat alleen maar ruis veroorzaakt. Misschien als je als adviseur uitgelegd krijgt wat je er mee kan dan is het handig.	[Background information to understand the tool is desirable]
2	Uhm, ja je moet als adviseur een goede instructie van jou of van iemand anders die weet hoe het werkt.	Achtergrond-informatie of instructie over de werking van de tool lijkt wenselijk [Background information about the tool's working seems desirable]
2	je moet wel heel goed aan een werkgroep kunnen uitleggen waarom je een penalty van een bepaalde zwaarte of iets, zeg maar, toe..kent. Ehm… En goed je mannetje kunnen staan, zeg maar, om daar een discussie over aan te gaan. Want.. ik kan nu, als ik aan de werkgroep moet uitleggen waarom ik die penalty 1 score heb gegeven, kan ik dat niet, zeg maar, makkelijk aan ze uitleggen. Dus dat is denk ik wel een vereiste, dat je goed ehm… beslagen ten ijs komt en ik denk, als laatste, dat je aan de werkgroep alleen het simpele overzicht moet laten zien. Ehm. En dat we als adviseurs daar allemaal uniform in geïnstrueerd moeten worden van: dat je niet nog de situatie zou krijgen dat in het ene cluster penalty 1 wordt gekozen en in het andere cluster niet d.. zo'n penalty, terwijl je dezelfde situatie hebt.	Achtergrond-informatie over de werking lijkt wenselijk om beslagen ten ijs te komen bij de vergaderingen met deelnemers [Background information about the tool's working seems desirable to come prepared to participant-meetings]
4	Alleen dit was mij nog niet helemaal duidelijk, oh dat kun je natuurlijk nu niet zien, bij die penalty, waarom je dan voor 4 kiest, maar goed dat mag je misschien nu ook nog niet zeggen. Ehm, maar dat is misschien meer achtergrondinformatie dan dat je dat echt in zo'n formulier moet zetten.	Keuze voor welke mate aggression onduidelijk [Choice for which size of aggression is unclear]
4	Ehh… ik vond eigenlijk de sessie uiteindelijk best wel kort om nu goed in te kunnen schatten hoe het dan echt werkt en of ik dan echt dingen mis, ofzo. Snap je wat ik bedoel? (…) Ja, misschien dan toch iets meer van die achtergrondinformatie over die, die, dat penalty, die methode, ofzo?	Geen achtergrondinformatie gegeven [No background information provided]
4	Ehm, alleen dat je dat hier bij de penalty, dat je dan zelf iets in moest vullen dat wist ik dan ook niet, maar goed daar krijg je dan misschien ook instructie voor.	Achtergrond informatie over penalty (aggression) ontbreekt nog [Background information

		about the the penalty (aggression) is still missing]		
4	Ja wat ook natuurlijk, eigenlijk wat we al soort van benoemd hebben, dat iets meer achtergrond wilde hebben over die, dat penalty eh, methode, die drie methodes die daar staan, twee?	Meer achtergrond informatie over penalty methode [More background information about the penalty method]		
3	Ehm, sowiezo moet je natuurlijk weten wat we precies bedoelen met prioriteren. Ehm, en ehm… Je moet natuurlijk ook.. je moet eigenlijk ook de tool een keer hebben gezien voordat je gaat prioriteren. (…) Dus eigenlijk is het wel handig dat je een keer het gezien hebt of een keer er in gespeeld hebt, zeg maar, voordat je daadwerkelijk met het prioriteren aan de slag gaat, van start gaat. (…) Ik denk dat de allerbeste methode is dat je er zelf een keertje wat mee hebt gedaan. Ehm, maar een demonstratie zou wel helpen.	Enige achtergrond lijkt noodzakelijk voordat je de tool daadwerkelijk in gaat zetten [Some background seems necessary when actually using the tool]		
1	Ik denk een handleiding, inderdaad. (…) Dat hoeft niet in de tool om het onnodig ingewikkeld te maken.	Achtergrond-informatie niet in de tool maar in een handleiding [Provide background information in a manual and not in the tool]	Background information can be provided in a manual, not necessarily within the tool itself	We will focus on providing additional information within an accompanying document, rather than extending the guiding information within the tool.
1	Ik denk dat het goed is en ook de kracht is van de tool nu en ik zou dit soort informatie er dan inderdaad niet bij zetten.	Achtergrond-informatie niet in de tool maar in een handleiding [Provide background-information in a manual and not in the tool]		
1	Ja… dan zou je bijna met tekstballonnetjes moeten werken met: stap 1 vul hier wat in, ofzo. Maar ik weet niet of dat mogelijk is en… in principe als je gewoon uitleg er bij hebt is het ook niet nodig.	Achtergrond-informatie in een apart document lijkt voldoende [Background information in a manual may be sufficient]		
1	Het lijkt in het begin als je ‘m ziet heel overweldigend, met alle valkjes en dingetjes. Maar als je ‘m eenmaal gaat gebruiken valt het reuze mee.	De tool lijkt relatief gebruiksvriendelijk na de eerste indruk en	It seems dificult to grasp the tool at first, but seems to	We understand that the first impression of the tool and its structure is complex. We hope

		gebruik [The tool seems relatively user-friendly after the first impression and use]	get more structured/ user-friendly after it is used and some background information is received	by providing additional background information about the structure and workings of the tool that the tool is immediately usable.
2	ik denk, zeg maar, als je als adviseur snapt hoe die werkt dat het dan, dat het dan een overzichtlijke en goed gestructureerde tool is. Maar in eerste instantie is het denk ik best wel complex om eh.. om te vatten.	Overzichtelijk en gestructureerd als de gebruiker de tool kent [Clear and structured tool once the user understands the tool]		
1	Ik vind hem best overzichtelijk, maar het is wel dat je 'm even moet gebruiken voor je door hebt hoe 'ie werkt. Het is niet dat je in een oogopslag denk van: oke, dit moet ik nu doen.	De tool lijkt relatief overzichtelijk, maar niet in eerste oogopslag [The tool seems relatively clear, but not in the first impression]		
1	Nee, want ik het begin ging ik de mist in met die ehh. Ranking. Dat ik hier de scores ging invullen. Later, als je dat zegt: ohja hier staat scores. Dan denk ik ja, dan is het ook dom dat ik het daar invul, maar… Automatisch begint je ook bovenaan in de tool en niet onderaan.	Structuur van de tool is niet meteen helder [Structure of the tool is not immedialtely clear]		
3	Ja ik vind het best logisch op elkaar aangrijpen. Ehm, misschien komt dat wel omdat ik een beetje weet wat de bedoeling is van de tool en waar het heen moet. Maar eh ik vond het best wel intuïtief werken, eigenlijk.	Gebruik van tool lijkt intuitief met enige achtergrond [Use of the tool seems intuitive when provided some background]		
4	Ja, nadat je mij zeg maar, nou, een paar keer geholpen had werd het duidelijk. Ehm, maar goed, we hebben het natuurlijk, over die scores hadden we het natuurlijk al een keer gehad he, dat ik, dat het voor mij niet helemaal overzichtelijk was, dat je, hoe je dan eh…	Structuur van de duidelijker na enkele aanwijzingen [Structure of the tool is more clear after some instruction]		
1	ja er zijn best wel beperkte mogelijkheden om te klikken en dingen in te vullen, dat werkt wel prettig. En… Ja het is fijn dat er dus dingen zwart zijn, dingen grijs. Dat je gewoon weet waar je wel iets mee kan doen en waar je niks mee kan doen.	Lay-out werk sturend [lay-out provides guidance for its use]	Using colors and restricting cells may guide the user experience within	We will provide some information about resticting cells in worksheets in an accompanying document.

			the tool	
1	Ja die items erin gaat heel makkelijk, gewoon met knippen en plakken vanuit zo'n ander document.	Eenvoudige input naar de tool [Easy input into the tool]	Easy way of input and output to/from the tool	-
2	Ja, weetje, op zich wel prima, als je.. als je gewoon zo kan knippen en plakken uit die voorbeeldscenariotjes is denk ik super fijn dat dat zo werkt.	Manier van input en output is gebruiks-vriendelijk. [Manner of input and output is user-friendly]		
2	Het formatje, zeg maar, waarin je dat kan aanleveren, eh, kan je zo super makkelijk er zo in plakken, dus dat is ook niet echt moeilijk.	Manier van input naar de tool toe is eenvoudig [Easy input into the tool]		
3	Hmm, ik vind sowiezo dat er een paar handige functionaliteiten zijn toegevoegd. Eh, zoals dat je makkelijk kan copy-pasten d'r uit. Eh, dat je ook kan copy-pasten vanuit een Word-bestand of een ander Excel-bestand naar de tool en dat het ook allemaal naar elkaar verwijst.	Manier van input naar en output van de tool is handig [Input and output is easy]		
4	nou, ik vond het best wel, op zich wel, redelijk duidelijk, alleen voor mij was het dan niet helemaal duidelijk nog hoe dat je dan de scores dan, zeg maar, er onder moest zetten. Ik dacht natuurlijk dat 'namen' dan de namen van de modules was, maar dat is dan de naam van de deelnemer eigenlijk.	Locatie van scores in tool kan onduidelijk zijn [Location fort he placement of scores in the tool might be unclear]	Score placement in the tool might be difficult te understand at first.	We hope by providing additional background information about the structure and workings of the tool that the tool is immediately usable.
1	Ik denk dat het goed is om de verenigingen in ieder geval te laten weten dat je, ehm, een soort penalty kan gaan toepassen. Ik weet niet of dat standaard altijd gedaan wordt of dat dat alleen wordt gedaan, zoals je zei, omdat er dan volgens belangen van de eigen vereniging wordt gestemd. Ik weet niet of dat makkelijk inzichtelijk is of dat altijd zo gebeurt. En.. Misschien is het goed om dan in ieder geval mensen van te voren op de hoogte te stellen van: mocht.. mochten ze dat doen, onbewust of bewust, dat we daar voor corrigeren. Aan de andere kant, misschien als je dat van tevoren al zegt gaan ze juist volgens de belangen stemmen, maar… Ik weet niet of dat erg is. [lacht]	Onduidelijk in hoeverre deelnemers inzicht moeten krijgen in de penalties en werking van de tool [It is unclear to what extent participant should have insights into the penalties and working of the tool]	Some instructions for participants seem appropriate.	Information about assigning priority scores are desirable. We provide additional information about the score (e.g. maximum score) and its interpretation in an additional document. However, we refrain from providing an example text for participants since we believe priority-setting is context dependent and instructions should be tailored accordingly.

1	Ik denk dat daar iets van achtergrondinformatie wel handig voor is. Dat ze straks niet verrast worden als ze gestemd hebben, van: ja maar we hebben allemaal vanuit de <*wetenschappelijke vereniging*> op die module gestemd, bij wijze van spreken ze zijn allemaal bij elkaar gaan zitten en hebben het samengedaan, en ze komen er niet uit? Ik denk dat dat wel een beetje voor onbegrip zorgt, als je dat niet uitlegt.	Enige achtergrond-informatie voor deelnemers lijkt wenselijk [Some background information for participants seems desirable]		
2	Ehm, nou ik denk dat 'ie heel complex is om aan werkgroepleden zonder uitleg te tonen, zeg maar, en om 'm te laten zien. Dus misschien dat je, als je dat zou willen doen, dat je dat, dat je daar een soort van nog gemakkelijker dashboard voor maakt of dat je echt alleen maar.. eh, dat kopietje van de ranking laat zien, maar niet het.. die hele eerste tablad, dat is denk ik veel te complex.	De tool zelf is te complex om aan deelnemers te laten zien zonder achtergrond-informatie [The tool seems too complex to show to participants without providing background information]		
2	Oke. Ehm, ja als ze hun scores gaan invullen hebben ze gewoon natuurlijk duidelijke instructies nodig en ik denk dat je ze vooraf wel iets moet vertellen over de, eh, penalties die je, zeg maar, kan toedienen. Niet… Ik weet niet zo goed of je daar strategisch stemmen nog vooraf mee kan beïnvloeden op een negatieve manier, maar.. Ik denk wel dat het goed is dat mensen vooraf geïnformeerd, zeg maar, worden over.. nouja gewoon, hoe het werkt en hoe ze het moeten invullen.	Duidelijke instructies voor deelnemers lijkt wenselijk [Clear instructions for participants seem desirable]		
2	Maar heel veel van die tekst die dan in het lichtgrij.. sorry, in het lichtgrijs staat, daar vraag ik mij van af van: ja wil je dat eigenlijk wel weten, of eh.. ja… anders denk ik altijd, zeg maar, zo van zo simpel mogelijk is toch zo mooi mogelijk dan.	Inhoud van de instructie/ tekst in de tool is misschien niet relevant [Contents of the instruction/tekst within the tool might not be relevant]	Contents of the instructions/texts within the tool may not be relavant to all users.	We keep the brief guiding/ informational texts within the tool, since it might help some users.
3	Verder vind ik het ook heel fijn dat er heel veel uitleg bij staat. Dus dat er precies bij staat wat het betekent en wat wat voor consequentie heeft. Dus dat als je bijvoorbeeld voor penalty-methode 2 kiest, dat je dat ook de penalty aggression moet invullen, maar dat als je voor penalty 1 kiest dat het dan niet zo is. Ja, dat vind ik heel verduidelijkend.	Voldoende info gegeven in de tool [Enough information is provided in the tool]	Contents of the instructions/texts within the tool may be sufficient for some users.	We keep the brief guiding/ informational texts within the tool, since it might help some users.
3	Ik denk zeker dat 'ie er voor zou kunnen corrigeren ja, ja. En ik denk ook zeker dat je zo een beteren spreiding krijgt wat welke WV belangrijk vindt. Maar ik denk dat het wel altijd nog belangrijk blijft om ook op inhoudelijke gronden te kijken. Dus ook als er zo'n penalty,	Uitkomst van de tool hoeft niet de uitkomst van de priority-setting assessment te zijn	The tool's output may not bet he outcome of the priority-setting	-

	bijvoorbeeld, over is gegaan, te blijven kijken: zijn de modules die nu zijn geprioriteerd, zijn dat nog steeds de relevante modules op inhoudelijke grond?	[Output of the tool is not necessarily the same as the outcome of the priority-setting assessment]	assessments and additional steps in the assessment may lead to a definitive outcome.	
4	Mean, variantie, standaardafwijking… Nee, ik dank dat dat op zich voor mij wel duidelijk.. dat je zo wel een goed idee krijgt van hoe de res.. of hoe noem je dat? Dat je de spreiding van de scores wel goed kan inschatten zo.	Heterogeniteits-analyse lijkt duidelijk te zijn [Heterogeneity analysis seems to be cleer	It seems clear how heterogeneity analyses work within the tool.	-
4	Alleen zat ik nog een beetje mee van: wil je, maar dat weet ik dus niet goed, wil je dan ook nog terug kunnen vinden uit welke richtlijn de module komt? Want dat kan je natuurlijk nu niet zo heel goed terugvinden. (…) Ja, nouja, som heb je best wel algemene titels ofzo. Bijvoorbeeld 'nazorg', ofzo, in noem maar wat. (…) Eh… En dan heb je later een module geprioriteerd, en dan moet je wel, hoe ga ik dan zelf weer terugvinden bij welke richtlijn die hoort? Dan moet je weer soort van naar je basis-bestand en dat daarin bekijken ofzo. (…) Ja, ik zit nou, ja richtlijn dubbelepunt modulenaam, dan wordt dit, of tenminste dit, dan wordt de eerste rij in die tabel natuurlijk best wel lang. Ehm, waardoor je dat wel minder goed kan lezen, dus dat zou ik denk ik niet willen, maar hmmm.. ja… ik zou het wel fijn vinden dat je het toch ergens, toch ergens in terug kan vinden. (…) Ja, of kan je dan niet bijvoorbeeld hier dat je dan zegt, jaa ik weet niet of daar ruimte voor is, dat je dan op die, in die eerste rij bij die prioriteringstool, dat je dan soort bolletje maakt ofzo, d'r boven, dat je dan kan zien welke richtlijn die modules horen?	Identificatie van de bron van de items ontbreekt [Identification of the source of the item is missing]	The tool is missing a function to identify the source (title) of an item in the tool.	We will not adjust the tool, but rather add guidance about source identification in a document accompanying the tool.
4	want als je dan zeg maar die gegeven krijgt, dan ga je als eerste waarschijnlijk de labels invullen. Dan zet je die daar na de RRV tabblad doet. Na die tool, zeg maar, ofzo. En dat je dan zegt van: het aantal stemmers, en dan hoe je het rank, ofzo.	Volgorde van input/ tabbladen inrichten [Order of tabs]	There might be a desire to change the order of the tabs within the tool according tot he priority-setting process.	Users may change the order of tabs according tot heir own use.
4	Maar, je moet er wel, denk ik, kijk, je moet er misschien wel best wel nog veel invullen. Maar geen 200, maar… 't zijn er misschien wel 50 ofzo.	Zorgen om hoeveelheid in te vullen ranks [Worries about the number of ranks to assign]	There might be worries that there are a lot of items to assign ranks to.	-

NB1: Two penalty methods were present in this preliminary version of the tool. In later versions we removed one of the methods, which also expired the need for using the tab ‘voters’ in the tool.
NB2: The remaining penalty-adjusted weighting method and penalty aggression parameter were later renamed to decay-adjusted weighting method and decay aggression, respectively.
NB3: Names of persons and organizations were anonymized and placed between angle brackets.

ADDITIONAL FILE 4

THE REPS-TOOL PROGRAMMED IN MICROSOFT EXCEL

The REPS-tool can be downloaded at https://doi.org/10.1371/journal.pone.0300619

ADDITIONAL FILE 5

RE-WEIGHTED PRIORITY-SETTING (REPS): A QUICK-START GUIDE TO THE REPS-TOOL

CONTENTS

1.
1. *Purpose of the REPS-tool*
2. *Role in the priority-setting assessment*
3. *Re-weighted range voting introduction*
4. *Tool structure*
5. *Input format*
6. *Obtaining priority scores*
7. *Data entry*
8. *Weighting methods and parameters*
9. *Assigning ranks*
10. *Heterogeneity analyses*
11. *Tool output*
12. *Adapting the REPS-tool*
13. *Reference list*

Note: the REPS-tool is currently programmed in Microsoft Excel and uses macros. Make sure that Microsoft Excel allows macros on your computer, otherwise the REPS-tool cannot be used. Microsoft Excel might also automatically block macros in Excel-files downloaded from the internet. To use such Excel-files, it is necessary to check the 'unblock' checkbox in the file properties of the Excel-file. To learn more, see: https://learn.microsoft.com/en-gb/DeployOffice/security/internet-macros-blocked

1. PURPOSE OF THE REPS TOOL

The main purpose of the REPS-tool is to aid in the priority-setting of any item (e.g. clinical practice guideline, systematic review, recommendation, key question, etc.) based on priority scores assigned by participants in the priority-setting assessment by providing a ranked list of items as its output. It aims to be a flexible tool that allows for all different kinds of procedures in the priority-setting assessment before and after using the tool, as long as the input complies to the input format requirement.

Two different types of weighting methods are available [1, 2]:

- The regular Re-weighted Range Voting method
- The decay-adjusted weighting method, based on the regular weighting method, where the decay pattern of individual weights can be adjusted

Note: the REPS-tool is not a consensus method. It ranks items based on priority scores assigned by a group of participants. Though, consensus might later be achieved using the tool's output by additional steps in the priority-setting assessment when desirable.

2. ROLE IN THE PRIORITY-SETTING ASSESSMENT

The REPS-tool is a function component in a priority-setting assessment. The function needs input according to a specific input format and produces an output in the form of a ranked list. The function's mechanism is re-weighted range voting to assign ranks.

All steps prior to using the function can be considered procedural steps ultimately leading to the correct input format for the function. Organization may wish to use their own set of selected priority indicators, may wish to ask participants to score multiple indicators per item and use the mean score, may wish to use mean scores of groups of stakeholders, and/or any other preferred procedural step. However, it should lead to data satisfying the input format for the REPS-tool.

Once data is entered in the REPS-tool and ranks are assigned, the tool's output can immediately be used as the outcome of the priority-setting assessment (Fig. 1). However, organizations may wish to add one or several procedural steps after the function component. For example, the tool's ranked top 15 output could be used to discuss and form a top 5 priority items as a definitive outcome of the priority-setting assessment.

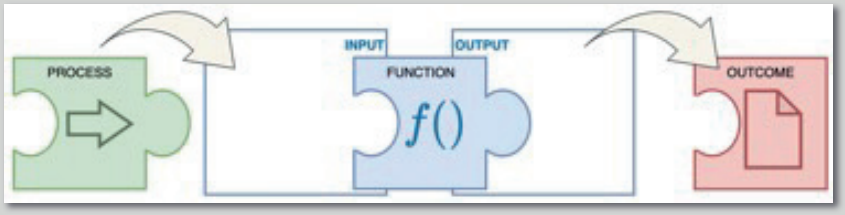


Figure 1 – *A process component leading to the input of the function component. The output of the function component is immediately used as the outcome of the priority-setting assessment, although the latter may not always desirable.*

3. RE-WEIGHTED RANGE VOTING INTRODUCTION

The REPS-tool uses re-weighted range voting as a mechanism to form a list of items ordered ascendingly by their assigned ranks. The ranks are assigned based on the highest sum score of the items. However, after every assignment of a rank to an item, a weight for each of the participants in the priority-setting assessment is being (re)calculated based on the scores that the participant already assigned to the previously ranked item(s). All participants have a weight starting at '1'. The higher the scores on the previous ranked items for each participant, the lower the individual

weight when recalculated. The lower the weight, the less influence on the next item to be ranked as the individual weights are multiplied by the participant's item scores. The re-weighted participants scores are summed to form new item sun scores and the highest item sum score is assigned the subsequent rank.

The calculation of the individual weights is, among others, based on the maximum scale score. Every participant scores items on a scale from 0 to a maximum predefined score (e.g. 5). There is no limit for items to receive the same priority score, so, in theory, a participant could rate all items with the same priority score.

The scale uses a distinctive description. It is not allowed to assign negative priority scores (e.g. -1). The scores can be interpreted as follows:

0 = No priority
1 = Lowest priority possible
…
Maximum scale score = Highest priority possible

The REPS-tool will not adjust the individual weight when a participant assigned a score of 0 or refrained from assigning a score to a winning item.

4. TOOL STRUCTURE

4.1 INTRODUCTION TO THE STRUCTURE

The REPS-tool consists of three worksheets: RRV (Fig. 2), Labels list (Fig. 3), and Ranking outcome (Fig. 4).

Entry of participant data and priority scores, heterogeneity analyses, and ranking is conducted in the 'RRV' worksheet. It contains six sections: item labels, ranking information, heterogeneity analysis, tool parameters, participants and scores, and ranking buttons.

4.2 SECTION A (FIG. 2A) – ITEM LABELS

Item names cannot be filled out in section A in the RRV work sheet. Rather, they are filled out under the Labels list worksheet and projected in section A (Fig. 2A). Item names of items to be prioritized (e.g. CPG names, section titles, recommendations, key questions, etc.) can be placed in the 'Labels list' worksheet (Fig. 3). If a list of item names is available, then the list can be copy/pasted into the Labels list worksheet (Fig. 3).

4.3 SECTION B (FIG. 2B) – RANKING INFORMATION

Here, (un)weigted sum scores can be seen per item in each respective column for informative reasons. A winner is automatically detected by the tool, showing a green cell with 'WINNER' in the winning item's column. Although assigning ranks to winning items can be

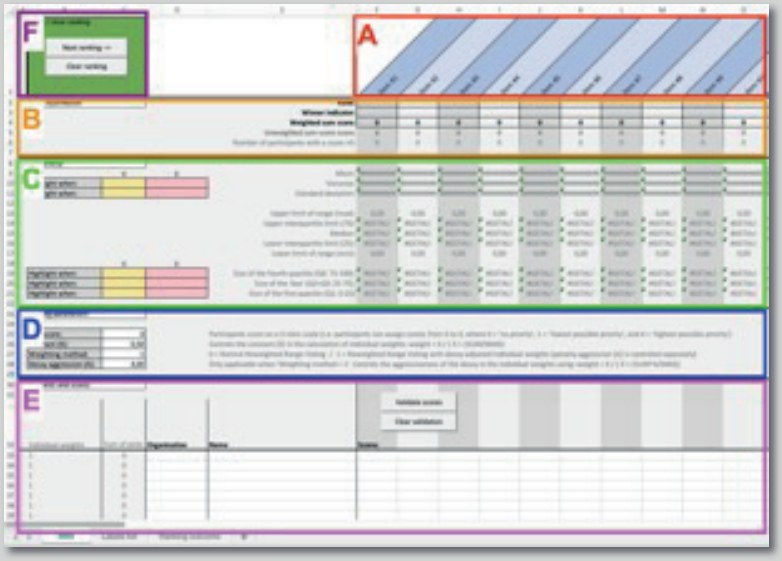
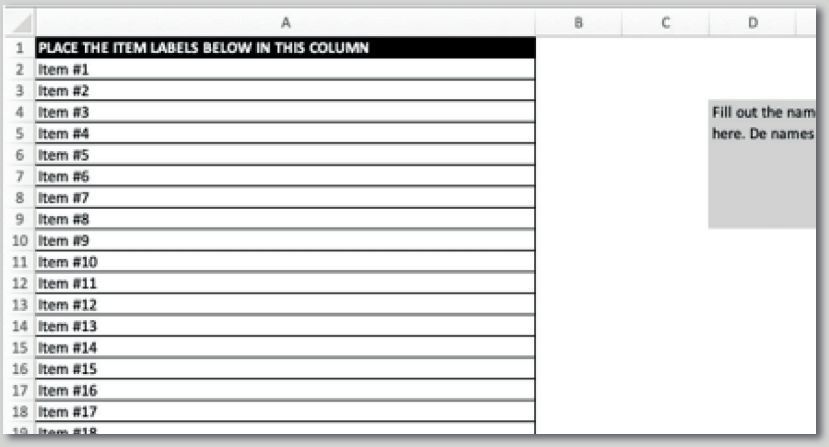


Figure 2 – *Overview of the 'RRV' worksheet in the REPS-tool opened in Microsoft Excel.*



	A
1	PLACE THE ITEM LABELS BELOW IN THIS COLUMN
2	Item #1
3	Item #2
4	Item #3
5	Item #4
6	Item #5
7	Item #6
8	Item #7
9	Item #8
10	Item #9
11	Item #10
12	Item #11
13	Item #12
14	Item #13
15	Item #14
16	Item #15
17	Item #16
18	Item #17

Figure 3 – *Detail of the 'Labels list' worksheet in the REPS-tool opened in Microsoft Excel.*

performed semi-automatically, a rank can also be assigned manually (e.g. in case of a tie) by inserting a number in the item's cell on the row named 'Rank' (i.e. immediately above the green cell indicating 'WINNER').

4.4 SECTION C (FIG. 2C) – HETEROGENEITY ANALYSIS

This section concerns heterogeneity analyses. For each item some measures of central tendency and measures of dispersion are displayed in the item's respective column.

4.5 SECTION D (FIG. 2D) – TOOL PARAMETERS

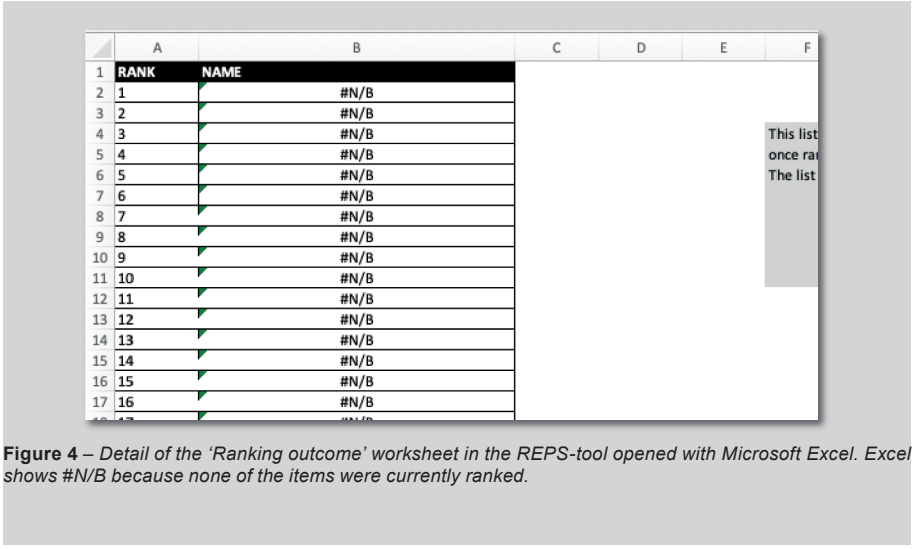
Tool parameters can be adjusted here. Available tool parameters are the constant, the maximum scale score, the weighting method, and the decay aggression (only available for the decay-adjusted weighting method). Changing these parameters will not result in the tool providing any direct feedback, but rather different formulas will operate in the background.

4.5 SECTION E (FIG. 2E) – PARTICIPANTS AND SCORES

The columns indicating the 'individual weight' and 'sum of picks' are informative for the user and are part of the calculations in the background. The column 'organization' and 'name' are optional, but can be used to identify the participant. Importantly, the field with 'scores' align with the items (columns) and participants (rows) and therefore forms a matrix.

4.6 SECTION F (FIG. 2F) – RANKING BUTTONS

This section contains two buttons: one to assign a rank to the identified winner and one to clear all assigned rankings. Items with assigned ranks are displayed in the Ranking outcome worksheet (Fig. 4). This list can be copy/pasted to word-processing software and is, effectively, the output of the REPS-tool.



	A	B	C	D	E	F
1	RANK	NAME				
2	1	#N/B				
3	2	#N/B				
4	3	#N/B				
5	4	#N/B				
6	5	#N/B				
7	6	#N/B				
8	7	#N/B				
9	8	#N/B				
10	9	#N/B				
11	10	#N/B				
12	11	#N/B				
13	12	#N/B				
14	13	#N/B				
15	14	#N/B				
16	15	#N/B				
17	16	#N/B				

This list
once ra
The list

Figure 4 – *Detail of the 'Ranking outcome' worksheet in the REPS-tool opened with Microsoft Excel. Excel shows #N/B because none of the items were currently ranked.*

Note: Worksheets can be protected (whether or not by password) so that cells and formulas are locked in their respective cells when desirable. Restricting cells may also result in users not able to select irrelevant cells guiding their experience in using the REPS-tool.

5. TOOL STRUCTURE

From Fig. 2A and Fig. 2E it could already be deduced that the columns represent the items to be prioritized and that the rows represent participants. The input format for the tool is thus a matrix with priority scores corresponding to an item and a participant:

Matrix 1 – *Input format for the REPS-tool.*

	Item 1	Item 2	Item 3	Item …
Participant A	Score 1,A	Score 2,A	Score 3,A	Score …,A
Participant B	Score 1,B	Score 2,B	Score 3,B	Score …,B
Participant C	Score 1,C	Score 2,C	Score 3,C	Score …,C
Participant D	Score 1,D	Score 2,D	Score 3,D	Score …,D
Participant E	Score 1,E	Score 2,E	Score 3,E	Score …,E
Participant…	Score 1,…	Score 2,…	Score 3,…	Score …,…

Any priority-setting process may precede, as long as the eventual data used for the REPS-tool complies to the input format in Matrix 1. For example, the scores in Matrix 2 (hypothetical example) are overall scores the participants assigned based on four priority indicators. However, it is also possible to provide the mean or median of separately scored priority indicators per item as input for the REPS-tool. Nonetheless, the options in the score scale should always include 0 (i.e. no priority), as a score of 0 will not adjust the individual weight.

Furthermore, when there are significant uneven numbers of delegates from participating stakeholders and this is considered undesirable, the mean or median of the participating delegates of stakeholders can be used as an input in the REPS-tool. Rows then become the stakeholder organizations or (sub-)groups instead of individuals (Matrix 3).

Matrix 2 – *Hypothetical example of a dataset (with a maximum scale score of 5) used for input in the REPS-tool to prioritize key questions for development.*

	What is the role of MR imaging in patients with a hepatocellular carcinoma?	What is the role of biopsy in the detection of a hepatocellular carcinoma?	What is the role of preopera-tive portal vein embolization in patients with a cholangio-carci-noma?	What initial approach should be used in patients with any biliary carci-noma: surgery or systemic therapy?
Oerbekke	4	4	4	5
Gaasterland	2	3	1	4
Van der Laan	0	4	4	2
Hooft	5	2	2	3

Matrix 3 – *Input format for the REPS-tool when using, for example, the mean of participating delegates per organization.*

	Item 1	Item 2	Item 3	Item …
Organization A	Mean 1,A	Mean 2,A	Mean 3,A	Mean …,A
Organization B	Mean 1,B	Mean 2,B	Mean 3,B	Mean …,B
Organization C	Mean 1,C	Mean 2,C	Mean 3,C	Mean …,C
Organization D	Mean 1,D	Mean 2,D	Mean 3,D	Mean …,D
Organization E	Mean 1,E	Mean 2,E	Mean 3,E	Mean …,E
Organization …	Mean 1,…	Mean 2,…	Mean 3,…	Mean …,…

6. OBTAINING PRIORITY SCORES

It is important to recognize that the REPS-tool uses a single score or no score (when participants refrained from assigning a score) per item. Thus, either the cell contains a number or is left blank. There are several ways to obtain a single score, for example:

- Ask to assign an overall priority score for each item
- Ask to assign a priority score on a single priority indicator for each item
- Ask to assign priority scores on multiple priority indicators for each item and average the assigned scores on the priority indicators for each item.

Organizations may have different needs and may operate in different contexts. It might be helpful to select priority indicators which seem relevant to the context and needs. For example, priority indicators might be dependent of the field the priority-assessment takes place, or for which purpose the item is prioritized for (e.g. update, development de novo, implementation), or even the health care system the organization is operating in.

Scores can be elicited by sending out an (online) survey, for example. In the future, priority-setting tools may be programmed into applications from where the scores might also be elicited.

Note: Literature reviews reporting priority indicators for guidelines are being published in scientific journals. Guideline developing organizations may use such overviews to select priority indicators relevant for their priority-setting assessment. Such literature includes:

- Martínez García L, Pardo-Hernandez H, Superchi C, Niño de Guzman E, Ballesteros M, Ibargoyen Roteta N, McFarlane E, Posso M, Roqué I Figuls M, Rotaeche Del Campo R, Sanabria AJ, Selva A, Solà I, Vernooij RWM, Alonso-Coello P. Methodological systematic review identifies major limitations in prioritization processes for updating. J Clin Epidemiol. 2017 Jun;86:11-24. doi: 10.1016/j.jclinepi.2017.05.008. Epub 2017 May 24. PMID: 28549931.
- El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. BMC Health Serv Res. 2019 Oct 15;19(1):692. doi: 10.1186/s12913-019-4567-2. PMID: 31615509; PMCID: PMC6792189.
- El-Harakeh A, Lotfi T, Ahmad A, Morsi RZ, Fadlallah R, Bou-Karroum L, Akl EA. The implementation of prioritization exercises in the development and update of health practice guidelines: A scoping review. PLoS One. 2020 Mar 20;15(3):e0229249. doi: 10.1371/journal.pone.0229249. PMID: 32196520; PMCID: PMC7083273.

7. OBTAINING PRIORITY SCORES

The item names in Matrix 2 are hypothetical examples of questions prioritized to be developed de novo. Item names are automatically displayed in the corresponding columns in the ‘RRV’ worksheet when the names are placed in the list in the ‘Labels list’ worksheet (Fig. 5).

Organization, participant names and scores are easily copied and pasted in the tool when using a dataset structured according to the input format. Names and priority scores from Matrix 2 were pasted in the REPS-tool (see Fig. 6). The item names can also be copy/pasted in the ‘Labels list’ worksheet. It is advisable not to cut/paste or move cells within the tool as underlying cell references may become dislocated, even when the worksheet is protected.

The columns ‘Name’ and ‘Organization’ do not necessarily have to contain data to perform the ranking of items in the REPS-tool. However, it is suggested to at least fill in the participant’s name (or identifier) when later identification is desirable (e.g. to check for errors in specific rows of data).

In this section the scores can be validated as well. That is, cells with input can be highlighted in color that contain text, values smaller than 0, or values larger than the maximum scale

score. Clicking on the button 'Validate scores' will highlight the cells (Fig. 7). The button 'Clear validation' will remove the highlights.

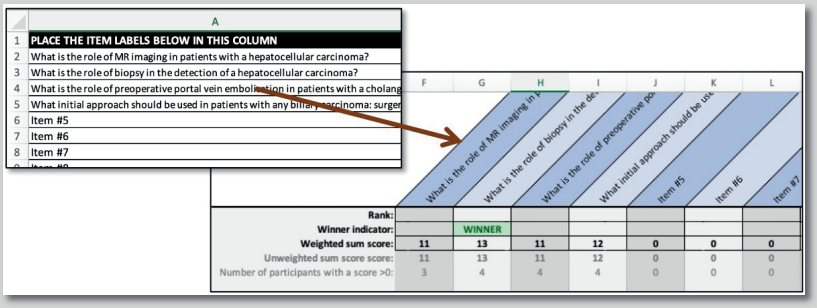


Figure 5 – *Item names placed in the 'Labels list' worksheet (Fig. 2) are automatically displayed in the 'RRV' worksheet (Fig. 1A)*

Note: Using the same order of participants and items from the dataset in the REPS-tool (i.e. in the 'RRV' and 'Labels list' worksheets) ensures that priority scores can be pasted in the tool without data entry errors. The item names may be horizontally oriented in the dataset. The transpose function Microsoft Excel (e.g. in a new workbook) can be used to create a vertically oriented list following the order of items used in the dataset. The vertical list can be pasted in the 'Labels list' worksheet and item names will appear automatically in the corresponding columns in the 'RRV' worksheet (Fig. 1A).

Note: When items have generic or similar names, consider adding identifiers to their labels. For example, when items are named 'Diagnosis of the primary tumor' and are sections from two different guidelines (e.g. biliary tract cancer and hepatocellular cancer):

- Diagnosis of the primary tumor [BTC]
- Diagnosis of the primary tumor [HHC]

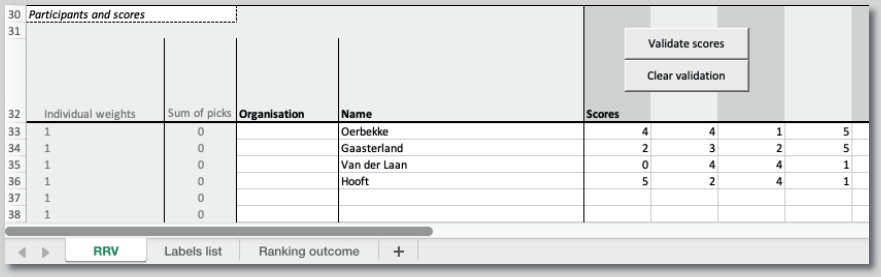


Figure 6 – *Names and scores were entered into the REPS-tool ('RRV' worksheet)*

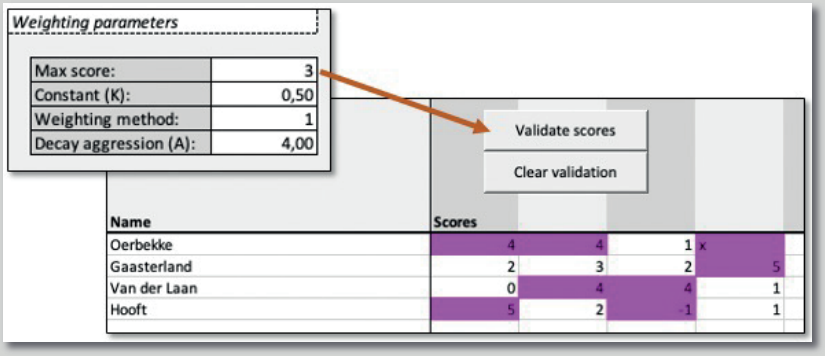


Figure 7 – *With a maximum scale score of '3' the 'Validate scores' button will highlight text, any score above 3, and any score below 0. The 'Clear validation' button will remove all highlights.*

8. WEIGHTING METHODS AND PARAMETERS

The REPS-tool has two weighting methods for individual weights: the regular re-weighted range voting method (Formula 1) and the decay-adjusted weighting method (Formula 2) [1, 2]. The latter formula can be used for disproportionate representation [2].

$$\textbf{(1)}\ individual\ weight = \frac{constant}{(constant + \frac{sum\ of\ participant's\ scores\ on\ ranked\ winners}{maximum\ scale\ score})}$$

$$\textbf{(2)}\ individual\ weight = \frac{constant}{(constant + \frac{sum\ of\ participant's\ scores\ on\ ranked\ winners^{A}}{maximum\ scale\ score})}$$

The regular re-weighted range voting method will result in a proportional representation (as assigned with priority scores) reflected in the ranking [1, 2]. The decay-adjusted weight (with A>1) results in a more aggressive decay of the individual weights (Fig. 8). A relative aggressive decay (e.g. A=4) causes a steep decline of the individual weight immediately with the first few points assigned to a winning item. This could theoretically leave room for boosting less represented perspectives in the ranking outcome, as participants 'having their way' with the first few winners will have a very low individual weight thereafter.

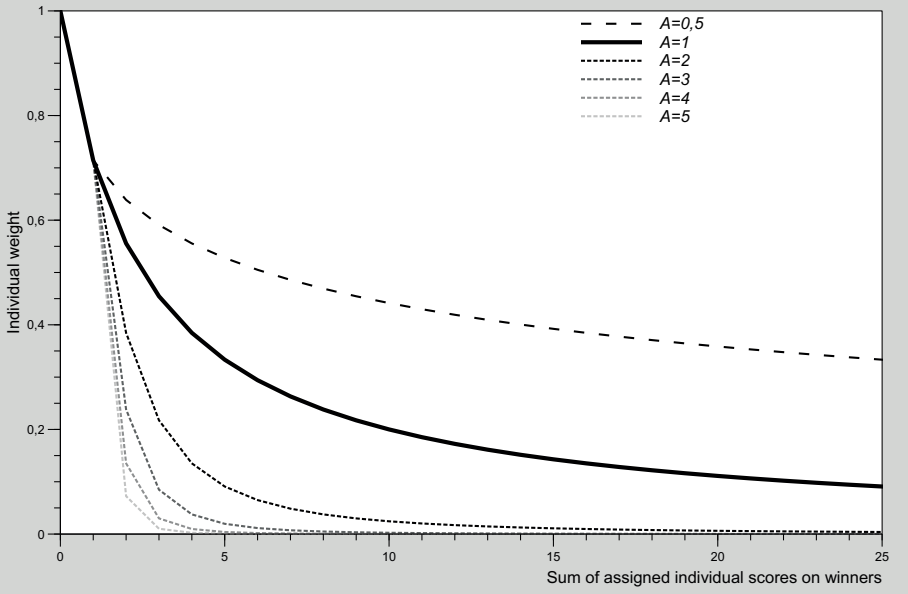


Figure 8 – *A graph showing the decay of individual weights using the decay-adjusted weighting method (Formula 2), a constant of K=0.5, and a maximum scale score of 10. The regular re-weighted range voting method (Formula 1) is identical to A=1.*

The weighting parameters of the individual weights can be selected in the weighting parameters section in the 'RRV' worksheet of the REPS-tool (Fig. 2D, Fig. 9). The value in the cell right to 'Max score:' indicates the maximum scale score. That is, the maximum score a participant is allowed to assign to an item. A brief explanation is provided on the right of the value in the REPS-tool which adapts according to the input value of the maximum scale score. For example, using a maximum scale score of 7, the participants essentially score on an 8-item scale ranging from 0 to 7.

The formulas to calculate individual weights need a constant in both the numerator and denominator in order to result in a weight of '1' when no items were ranked or when no scores were assigned to the previous ranked item(s). The constant can be any positive number [1].

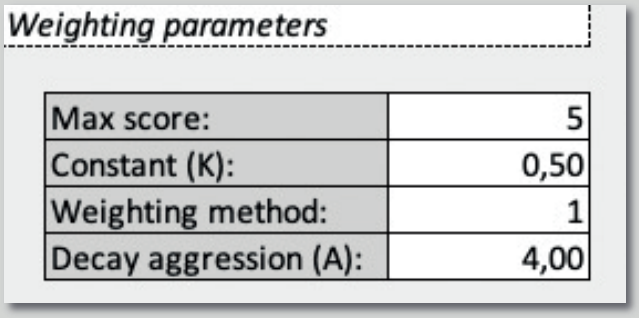


Figure 9 – *Weighting parameters in the 'RRV' worksheet.*

By changing the value of the cell right to 'Weighting method:', the weighting method can be changed. The regular re-weighted range voting method (Formula 1) is used when the value is '0', while the decay-adjusted method (Formula 2) is used when the value is '1'. The value next to the cell 'Decay aggression (A):' is the value of exponent A in Formula 2.

Note: The tool does not provide immediate feedback when changing parameters. It switches formulas in the background to calculate the individual weights, for example when switching between 0 and 1 as the weighting method.

9. ASSIGNING RANKS

The REPS-tool immediately indicates the first 'winner' in green when scores are entered (Fig. 10). Make sure not to assign ranks until all data required following the input format is entered to prevent a faulty ranking. In our hypothetical example from Matrix 2, the first winner is "What is the role of biopsy in the detection of a hepatocellular carcinoma?" because it had the highest sum score (i.e. 13).

By clicking on the 'Next ranking –>' button (Fig. 11), the identified 'winner' receives its rank and the next 'winner' is identified to receive a rank. The button enables a semi-automatic ranking of items (Fig. 12). A message will appear to inform the user there is no item left to rank when the button is clicked and there are no more items left. Ranks can also be manually entered in the row 'Rank:' (e.g. fill out '1' in the cell above the cell indicating the winner in Fig. 10). After accepting the value in the cell, the REPS-tool will immediately identify the next 'winner'.

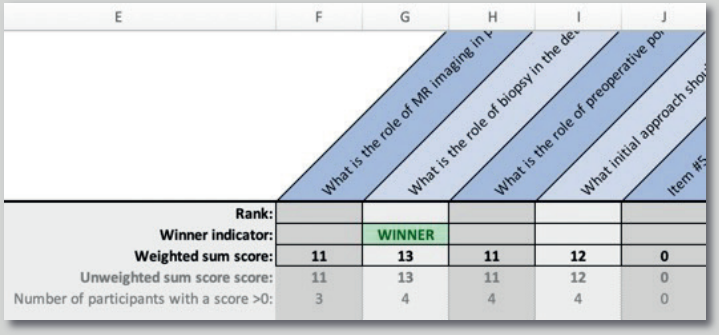


Figure 10 – *Automatic indication of the winner in the ranking information section (also see Fig. 2B). A rank is not yet assigned to the winner at this point.*

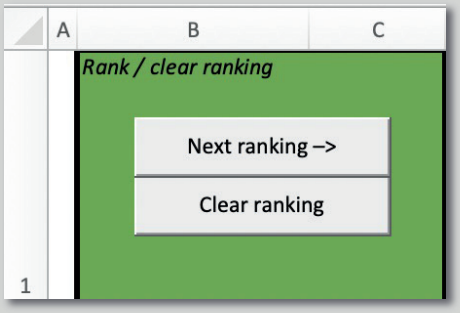


Figure 11 – *Buttons for the semi-automatic assignment of ranks and for clearing all ranks.*

The row 'Rank:', containing all of the assigned ranks to the items in the corresponding columns (Fig. 10), can be cleared at once by clicking on the 'Clear ranking' button (Fig. 11).

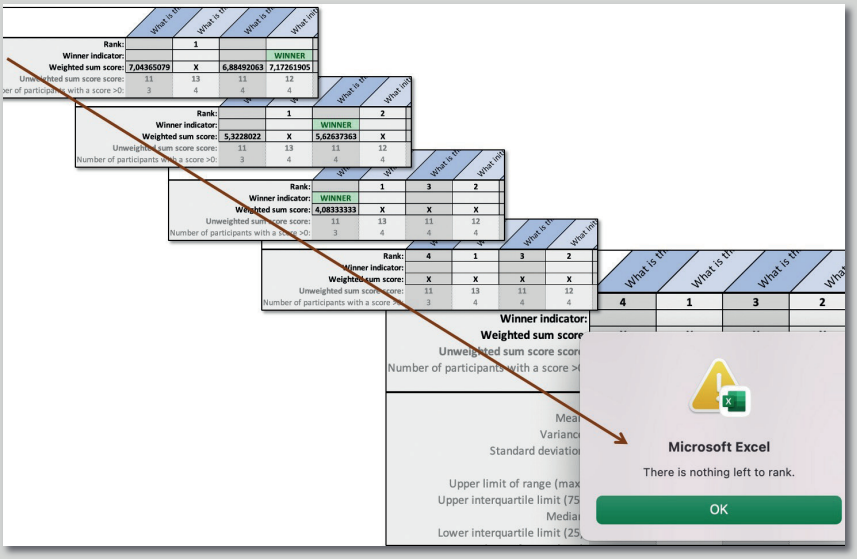


Figure 12 – *Semi-automatic ranking using the 'Rank next –>' button (Fig. 11) until there are no items left to rank, which prompts a message to inform the user.*

In some cases, the (new) sum scores of multiple items may be equal and these items are tied for a rank. Equal sum scores are most likely to happen during the first few assignments of ranks, whereafter it probably becomes less likely that items have exactly the same sum score due to the individual weighting of the assigned scores.

The REPS-tool will indicate which items have identical scores with 'MULTI (n)' in red when tied for a rank (Fig. 13). Here, n displays the number of items with identical scores. Clicking the 'Next ranking –>' button when multi-winners are identified will not assign a rank to any of these items. The 'Next ranking –>' button, however, prompts a message indicating for which rank the items are tied. This rank needs to be manually assigned to one of the multi-winner items. There are no prespecified rules for this decision yet, however heterogeneity analysis may aid in the

decision. For example, one might choose an item where there is some agreement among the participants reflected by a smaller variance, standard deviation, or range compared to the other multi-winner item(s).

All items who received a rank will be displayed in the Ranking outcome worksheet ordered by their rank. When deciding to change the tool's parameters after a ranking is completed, the tool does not refresh the assigned ranks automatically. Thus, after changing the tool's parameters, the ranking should be cleared before new ranks are assigned according to the new parameters.

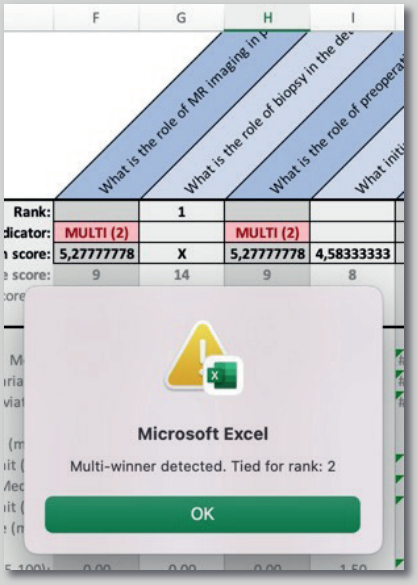


Figure 13 – *Two items are tied for a rank as indicated by MULTI(2) in the tool. The REPS-tool identifies which items are tied by displaying MULTI(n) in the item column and shows how many items are tied within the parentheses. Clicking the 'Rank next –>' button (Fig. 11) will prompt a message showing for which rank the items are tied. In this case there is a tie for rank 2.*

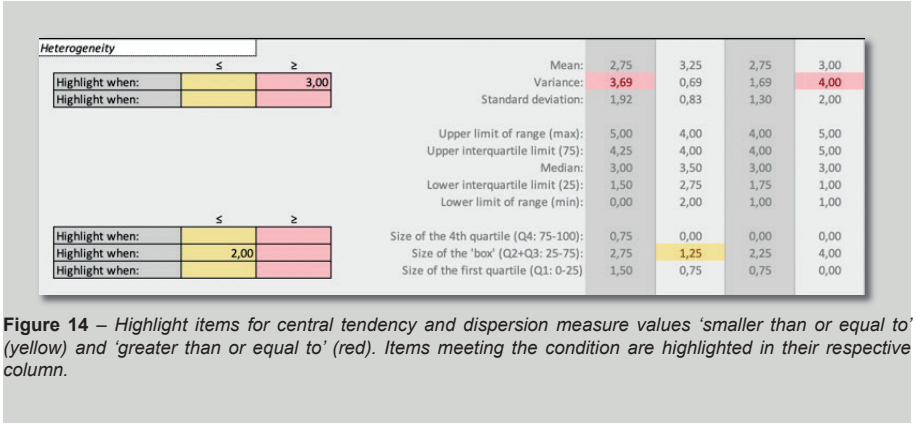
Note: Do not assign identical ranks to multiple items. These items will not be displayed correctly in the 'Ranking outcome' worksheet. For example, do not manually assign the same rank to the items identified as a multi-winner.

10. HETEROGENEITY ANALYSES

The REPS-tool has a section for heterogeneity analyses (Fig. 2C). Analyses are carried out on the unweighted scores assigned by the participants. Central tendency measures (i.e. mean and median) are provided, accompanied by measures of dispersion (i.e. variance, standard deviation, range, and interquartile range) and sizes of the quartiles.

Items meeting specific conditions, such as a variance greater or equal to three, can be highlighted in the REPS-tool (Fig. 14). Fill out and accept a value in the yellow (smaller or equal to) or red (greater or equal to) cell right to 'Highlight when:' for the tool to automatically highlight items meeting the condition.

When items are tied for a rank and the rank has to be entered manually, heterogeneity analyses may be of use in deciding which item receives the rank. Although there is currently no consensus and there are no predefined decision-rules, the measures of dispersion might indicate the level agreement among participants. For example, large variance could mean that the participants do not agree about the priority while little variance could indicate that there is more agreement about the level of priority. One might choose, for example, to assign the rank to the item with the smallest variance.

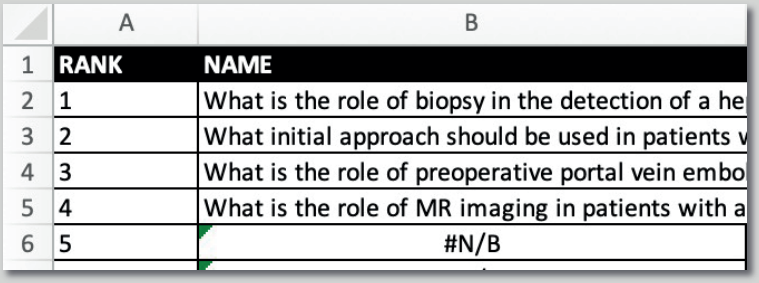


Heterogeneity	≤	≥					
Highlight when:		3,00	Mean:	2,75	3,25	2,75	3,00
Highlight when:			Variance:	3,69	0,69	1,69	4,00
			Standard deviation:	1,92	0,83	1,30	2,00
			Upper limit of range (max):	5,00	4,00	4,00	5,00
			Upper interquartile limit (75):	4,25	4,00	4,00	5,00
			Median:	3,00	3,50	3,00	3,00
			Lower interquartile limit (25):	1,50	2,75	1,75	1,00
			Lower limit of range (min):	0,00	2,00	1,00	1,00
	≤	≥					
Highlight when:			Size of the 4th quartile (Q4: 75-100):	0,75	0,00	0,00	0,00
Highlight when:	2,00		Size of the 'box' (Q2+Q3: 25-75):	2,75	1,25	2,25	4,00
Highlight when:			Size of the first quartile (Q1: 0-25)	1,50	0,75	0,75	0,00

Figure 14 – *Highlight items for central tendency and dispersion measure values 'smaller than or equal to' (yellow) and 'greater than or equal to' (red). Items meeting the condition are highlighted in their respective column.*

11. TOOL OUTPUT

The output of the tool is a list of items ascendingly ordered by the assigned ranks to the items. Only items that were (manually or semi-automatically) assigned a rank to in the 'RRV' worksheet are displayed in the 'Ranking outcome' worksheet of the REPS-tool (Fig. 15). It is possible to create a top 10 or a top 15, by (manually or semi-automatically) assigning a rank to the first 10 or 15 winners, respectively. That means that solely items that had a rank assigned to them in the 'RRV' worksheet are ordered and displayed in list in the 'Ranking outcome' worksheet. The list can be copied from the worksheet to other applications, such as Microsoft Word.



	A	B
1	RANK	NAME
2	1	What is the role of biopsy in the detection of a he
3	2	What initial approach should be used in patients v
4	3	What is the role of preoperative portal vein embo
5	4	What is the role of MR imaging in patients with a
6	5	#N/B

Figure 15 – *The list of items that received a ranking ordered by rank in the 'Ranking outcome' worksheet. Microsoft Excel shows #N/B because a fifth item was not assigned a rank.*

12. ADAPTING THE REPS-TOOL

The formulas in the worksheets and the worksheets themselves are not locked in the REPS-tool. Therefore, the tool can be programmed in Microsoft Excel according to specific needs not currently programmed in the tool. For example, it is conceivable to wish for a different weighting of individual weights for patients participating in the priority-setting assessment. Or, a different set of weighting methods could be programmed than currently available in the tool. Although these aspects are not features in the current version of the REPS-tool, it should be programmable using Microsoft Excel when desirable. The REPS-tool may be further edited by programming specific wishes for an organization in the tool. In fact, we would encourage any adaptations that would allow for its use for different contexts and needs. This includes priority-setting outside the context of clinical practice guidelines.

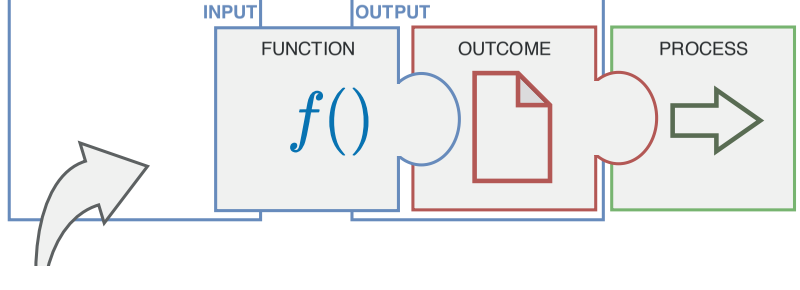
13. REFERENCE LIST

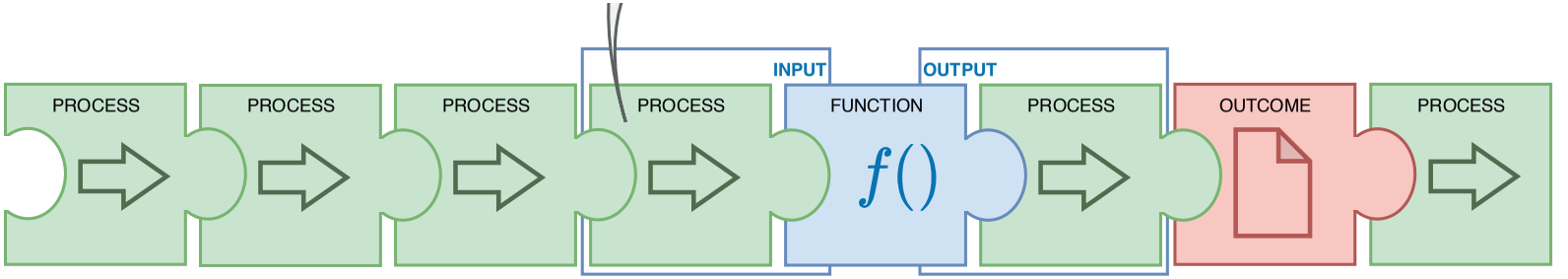
1. Kok J, Smith WD. Re-weigted Range Voting - a Proportional Representation voting method that deels like range voting [cited 2023 15th of May]. Available from: https://www.rangevoting.org/RRV.html.
2. Smith WD. Reweighted range voting – new multiwinner voting method. 2005.

ADDITIONAL FILE 6

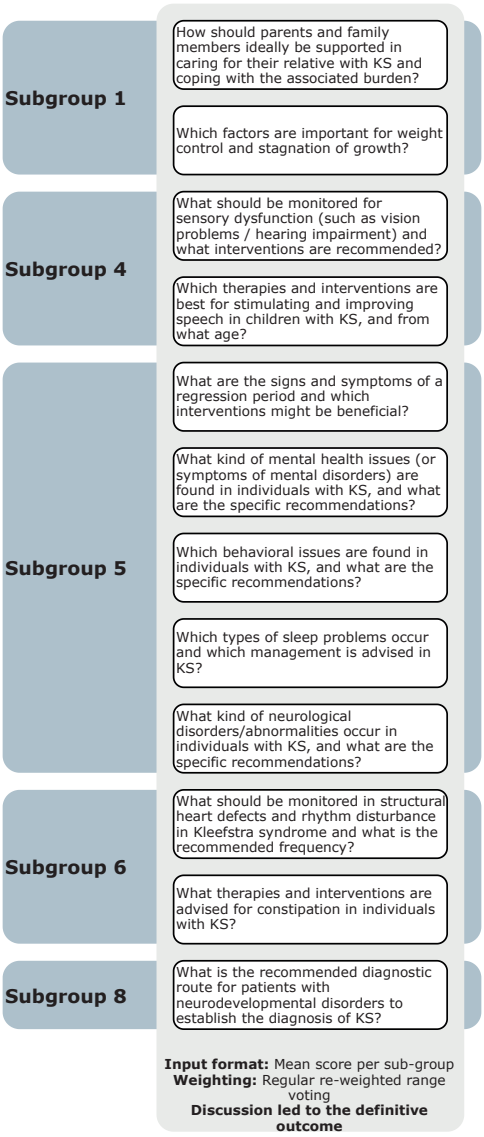
PRIORITY-SETTING IN THE KLEEFSTRA SYNDROME GUIDELINE

Calculation and ranking *Input:* The sub-groups' mean scores per key question / the participants' individual scores per key question in a matrix. *Function:* Regular re-weighted range voting / decay-adjusted voting (A=2 / A=4). *Output:* Four ranked list of key questions.	Four ranked top 10 lists obtained under four conditions in the function: *- mean subgroup score and regular re-weighted range voting* *- individual scores and regular re-weighted range voting* *- individual scores and mild decay-adjusted voting (A=2)* *- individual scores and aggressive decay-adjusted voting (A=4)*	**Evaluation** Panel members (including patient representatives) were asked in an online poll which of the four hypothetical top 10's they preferred.

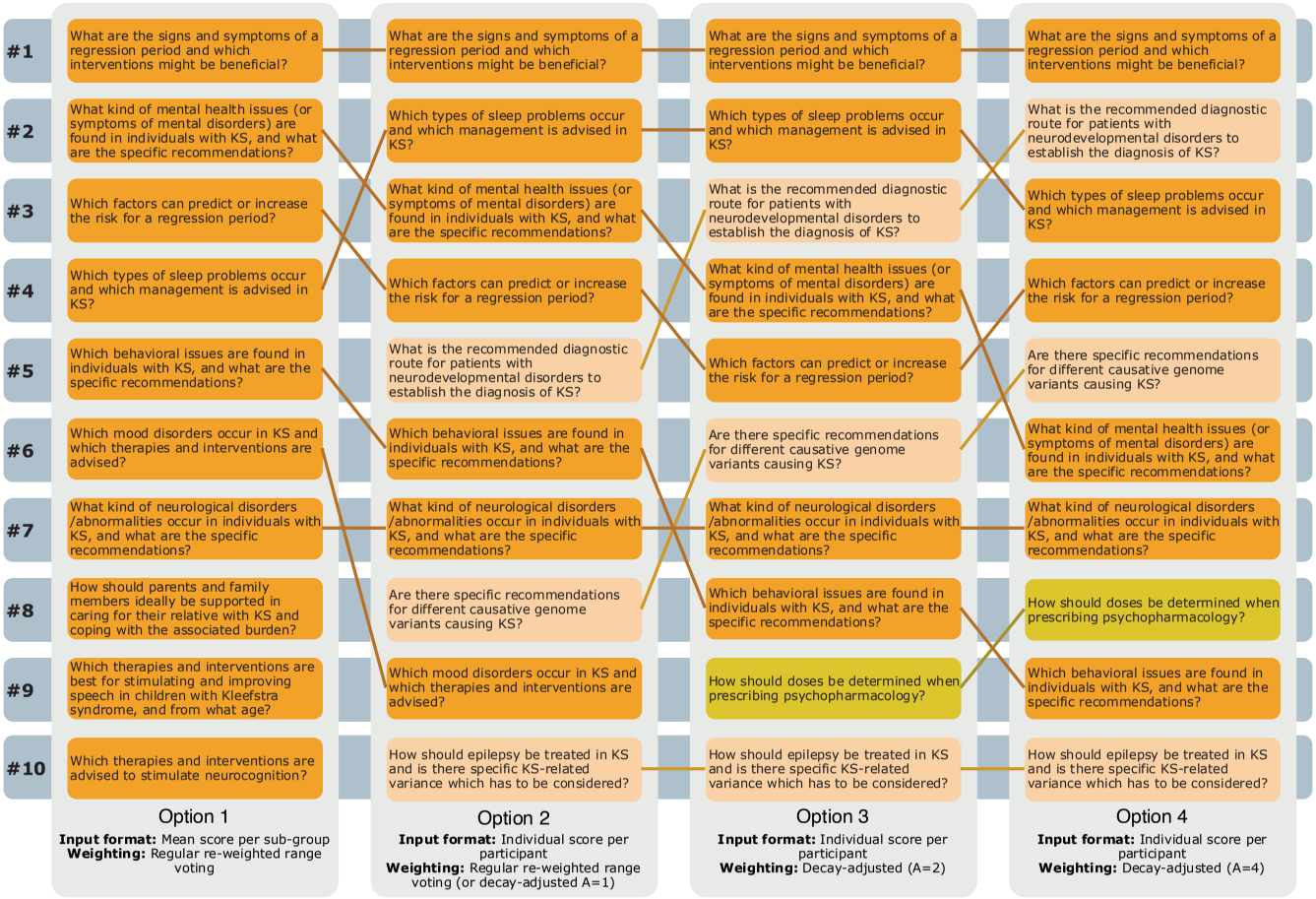




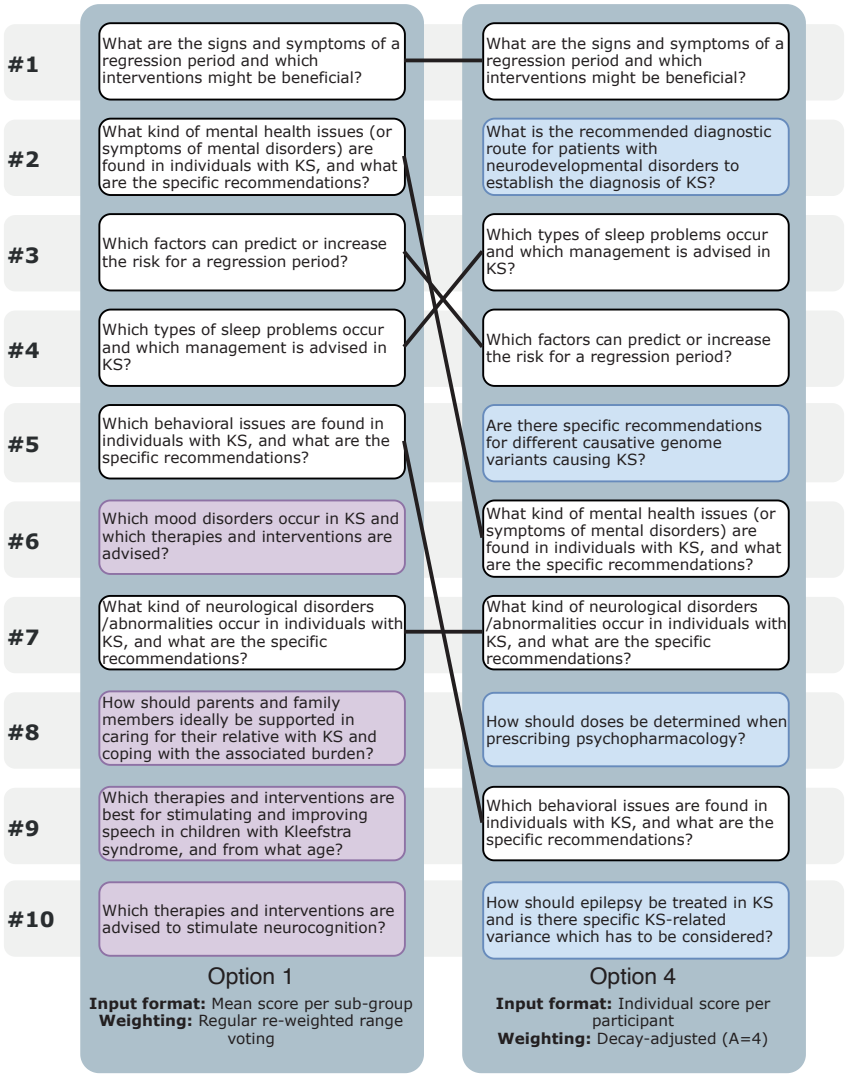
A6 Figure 1. The priority-setting assessment in the Kleefstra syndrome guideline. *The priority-setting assessment (colored row) was used to prioritize 12 key questions for development in the Kleefstra syndrome guideline (see S6 Fig 2). The parallel assessment (grey row with colored outline) shows how input from the priority-setting assessment was used to test different four conditions in the REPS-tool (see S6 Fig 3), whereafter the Kleefstra guideline panel were asked which output they preferred.*



A6 Figure 2. The final outcome of the priority-setting outcome in the Kleefstra syndrome guideline. *A top-18 output of the REPS-tool was presented to the core group and discussions led to the final selection of 12 key question for development in the Kleefstra syndrome guideline. The figure shows which input and weighting method was used and which key questions were ultimately selected per subgroup in the guideline panel through discussing the REPS-tool's output in the core group.*



A6 Figure 3. Four ranked top-10 outputs resulting from four conditions in the tool using input from the priority-setting assessment in the Kleefstra guideline. *The figure displays the four hypothetical options the Kleefstra syndrome guideline panel could indicate they preferred. Reading the figure from left to right: introduction of a new color in a column (i.e. light orange and mustard yellow) indicates the appearance of a key question now ranked in the top-10 which was not ranked in previous options. The crosslinks show how questions changed in ranking depending on the input format and weighting method.*



A6 Figure 4. Comparing the two hypothetical top-10 options most preferred by the Kleefstra syndrome guideline panel. *The figure shows two ranked top-10s obtained using different input formats and weighting methods. The purple highlighted boxes in Option 1 are key questions that did not appear in Option 4. The blue highlighted boxes in Option 4 are key questions that did not appear in Option 1. Six key questions appeared in both Option 1 and 4, with crosslinks showing the change in rankings between options. Note that Option 4 resulted in a #2 ranking of a question that did not appear in Option 1.*

ADDITIONAL FILE 7

THE SCORES OF ALL INDIVIDUALS PARTICIPATING IN THE KLEEFSTRA PRIORITY-SETTING GUIDELINE

The matrix can be downloaded at https://doi.org/10.1371/journal.pone.0300619

CHAPTER 7

Exploring transparent reporting and data availability of systematic reviews to identify subgroup evidence: imaging for suspected hepatocellular carcinoma in the non-cirrhotic liver

M.S. Oerbekke
R.A. de Man
F.G.I. van Vilsteren
M.W. Nijkamp
E. Tjwa
C.M.W. Gaasterland
M.J. van der Laan
L. Hooft

Submitted (letter to the editor)

ABSTRACT

We aim to illustrate the role of complete and transparent reporting coupled with access to data sourced from published systematic reviews, especially assisting in the identification of evidence for subgroups within the context of a rare disease. To accomplish this principle, we provide a real-world example encountered during the revision of the Dutch clinical practice guideline for hepatocellular carcinoma. Specifically, we retrieved insights from two Cochrane reviews to identify direct evidence concerning the diagnostic test accuracy of computed tomography and magnetic resonance imaging for detecting hepatocellular carcinomas in suspected patients without liver cirrhosis. Through reusing the Cochrane review authors' efforts already undertaken in their exhaustive literature search and selection, we successfully identified relevant direct evidence for this subgroup of suspected patients without cirrhosis and performed an evidence synthesis within the constraints of limited resources for the guideline revision. This approach holds the potential for replication in other subgroups in the context of rare diseases, contingent on the transparent and complete reporting of systematic reviews, as well as the availability and accessibility of their extracted data. Consequently, we underscore the importance of adhering to established reporting guidelines for systematic reviews, while simultaneously advocating for increased availability and accessibility to data. Such practices would not only increase the transparency and reproducibility of systematic reviews but could also increase reusability of their data. In turn, the increased reusability could result in reduced resource utilization in other sectors such as the guideline developing community as we show in our example.

Keywords: Systematic reviews, clinical practice guidelines, rare diseases, subgroups, reporting guidelines, data availability, data accessibility, data reusability, resource utilization

TO THE EDITOR

In 2022 the incidence of patients with hepatocellular carcinoma (HCC) in the Netherlands was n=649 (502 males, 147 females) with a five-year prevalence of n=1303, based on preliminary data in the Netherlands Cancer Registry.[1] This qualifies as a rare disease, given that the population size in the Netherlands was N=17.815.508 as of December 2022.[2] Liver cirrhosis is usually observed in patients with HCC, although the disease may also develop without cirrhosis in an even smaller group of patients. This may vary across populations, where 12%[3] and 26%[4] of the patients with HCC did not have liver cirrhosis within studies in the United States and Taiwan, respectively. Diagnostic imaging indicators of HCC in patients without liver cirrhosis have a lower specificity compared to those in patients with cirrhosis.[5] International clinical practice guidelines (CPGs) strongly recommend a pathology-confirmed diagnosis for the subgroup without cirrhosis,[5, 6] a practice that appears to be comparatively more prevalent among these patients.[3, 4]

Recently, there has been a revision of the Dutch HCC diagnosis and treatment CPG from its prior publication in 2013, which currently is awaiting endorsement in its finalization phase. The multidisciplinary CPG panel wished to provide guidance regarding the utilization of solely computed tomography (CT) or magnetic resonance imaging (MRI) for diagnosing HCCs in suspected patients without liver cirrhosis, as an alternative to a pathology-based diagnosis. We developed this guidance as a dedicated segment within the CPG, supported by a comprehensive synthesis of the evidence. However, a literature specialist demonstrated that even a specific search strategy for our evidence synthesis would result in a greater number of primary studies (i.e. n=1813 from Embase and n=1383 from Ovid/MEDLINE, duplicates not removed) than our capacity allowed for processing within a single CPG segment when considering the available resources for the CPG revision. Consequently, an alternative search strategy was developed and centered around published systematic reviews (SRs), resulting in n=33 hits after removing duplicates. The search strings for Embase and Ovid/MEDLINE involved keywords for HCC, CT, MRI, histology, diagnostic accuracy, and SRs and will be available upon request via www.richtlijnendatabase.nl once the revised guideline achieves endorsement and is published. One of the involved CPG panel members selected 10 potentially relevant SRs based on their titles and abstracts (TIAB), including two Cochrane Diagnostic Test Accuracy reviews.[7, 8] The supporting CPG methodologist read these 10 SRs in full-text and determined with the involved CPG panel members that none of these SRs reported relevant and direct information for the subsample of patients without cirrhosis. Therefore, none of the SRs could be directly used as an evidence synthesis in the CPG segment for the Dutch HCC CPG.

Both Cochrane reviews, however, reported the number of included studies that had 100% prevalence of cirrhosis and conducted a sensitivity analysis for studies with a prevalence of >90% versus studies <90% prevalence.[7, 8] This indicated that the Cochrane reviews captured primary studies that also recruited suspected patients without liver cirrhosis. These studies could possibly report (sub)analyses for the group of patients without cirrhosis. The Cochrane reviews transparently reported their search strategy, including sensitive search strings. The CPG methodologist judged that these search strategies (see Table 1) indeed could encase the retrieval of studies with a (sub)sample of patients without liver cirrhosis.

We reused the potentially relevant hits from both Cochrane reviews, comprising of the included studies and the referenced excluded studies. For the included studies,

Table 1 - Summary of the search and retrieval of the two Cochrane diagnostic test accuracy reviews.

Author, year	Modality	Generalized search string*	Search retrieval after removing duplicates (n)	TIAB potentially relevant (n)	TIAB excluded (n)	Full-text included (n)	Full-text excluded (n)
Nadarevic 2021 [8]	CT	(CT) AND (HCC) AND (liver disease)	25230	165	25065	21	144
Nadarevic 2022 [7]	MRI	(MRI) AND (HCC) AND (liver disease OR cirrhosis)	9661	117	9544	34	83

*This is a generalized version of the search sting without the synonyms the authors have used. There are differences within each search string depending on which database was sought. The complete search strings are available in the original publication of the Cochrane Diagnostic Test Accuracy reviews.
CT: Computed Tomography, HCC: Hepatocellular Carcinoma, MRI: Magnetic Resonance Imaging, n: number of studies, TIAB: Title/abstract

we downloaded the Cochrane reviews' data files from the Cochrane Library. These files contained data about the prevalence of cirrhosis in the included studies. We selected those studies which had a prevalence of <100% or when the prevalence was not reported (n=12), since these could potentially report (sub)analyses for patients without cirrhosis. The CPG methodologist read these papers full text and determined with the involved GPC panel members that two studies[9, 10] met the inclusion criteria for our evidence synthesis. Given that the full-text excluded studies from the Cochrane reviews were explicitly cited, we were able to compile and access this set of n=214 studies (duplicates removed). One CPG panel member screened these primary studies on TIAB and identified a set of n=18 potentially relevant studies for our evidence synthesis. The CPG methodologist read these 18 studies full-text and determined with the involved CPG panel members that one additional study[11] met the inclusion criteria. Overall, three studies met the inclusion criteria of our evidence synthesis for uptake in the GCP segment for the Dutch HCC CPG revision. The process is visualized in Figure 1.

Through this method, we effectively have reused the Cochrane review authors' efforts in the literature search, the literature screening, and the data-extraction to identify direct evidence for the test accuracy of CT and MRI in the subgroup of patients without liver cirrhosis. This was only possible because of the transparent reporting of information by the Cochrane review authors and the availability and accessibility of data within both reviews. It is important to note that this method does not extend to identifying primary studies published after the search dates reported in the Cochrane reviews. Nevertheless, it provided a systematic and time-efficient strategy for searching and screening potentially relevant

literature up to the Cochrane reviews' search dates. The TIABs of 33 systematic reviews and 214 primary studies were screened instead of n=25230[8] and n=9661[7] primary studies for CT and MRI, respectively. With an estimate that manually screening 60 to 120 TIABs takes approximately an hour,[12] screening the Cochrane reviews' search retrieval would take about 210-420 hours for CT and 80-161 hours for MRI (not regarding potential duplicates). Our own initial search strategy for primary studies would have taken about 15-30 hours to screen on TIAB for the retrieval from Embase alone. However, through reusing the Cochrane authors' efforts, the involved CPG panel members invested far less time in screening the TIABs for our evidence synthesis in the CPG revision, which would be about 2-4 hours according to the estimate of 60 to 120 abstract per hour. The CPG segment (including its supporting evidence synthesis with data extraction table and exclusion table) will be published online at www.richtlijndatabase.nl once the revision is endorsed.

The European Association for the Study of the Liver[5] and European Society for Medical Oncology[13] CPGs for HCC from 2018 do not seem to reference any of these primary studies for the test accuracy of CT or MRI in patients without cirrhosis. The latest American Association for the Study of Liver Diseases CPG from 2023,[6] however, does reference one of the identified primary studies.[9] The American CPG referenced an additional primary study[14] which we had not identified through both Cochrane reviews. This particular study was published after the search date in one of the Cochrane reviews,[8] making it impossible to be included. However, it was published shortly before the last search date in the second Cochrane review.[7] The study might not yet have been indexed properly in the databases this short before the last search date in the Cochrane review, resulting in its omission. Alternatively, in both Cochrane reviews, the authors exclusively included studies that reported per-patient analyses because they were interested in the ability of CT and MRI to detect patients with HCC, while per-lesion analyses typically provides information about staging.[7, 8] The study might have been excluded during the TIAB screening phase based on its abstract, which suggested that a per-lesion analysis was used. The reported Ovid MEDLINE search string[7] is, however, capable of capturing the study as it was found when we reproduced the search in Ovid MEDLINE on the 24th of July 2023. In this situation, having access to the complete search retrieval would have been beneficial allowing us to check whether the study was absent in the retrieval altogether or was excluded in the TIAB screening phase.

With the hereby described process during the Dutch HCC CPG revision, we have shown how information and data in SRs was used by the CPG working group and methodologist to identify direct evidence for a subgroup in a rare disease while under resource constraints. The potential for such a process to be applicable to subgroups in other rare diseases depends on the transparent and complete reporting of information in SRs, alongside with the availability of extracted data. This is essential for two reasons. Firstly, it allows for the verification whether the search strategy could have captured the subgroup of interest, and to understand which and when databases were last searched. Secondly, having access to this information and data are necessary to identify and retrieve the relevant literature for your own evidence synthesis. We therefore stress the importance of adhering to reporting guidelines for SRs, such as the PRISMA statement[15] and its extensions. Moreover, we emphasize the value of providing extracted data from primary studies in a data file format that can be accessed using widely available or free software. Although authors generally acknowledge the importance

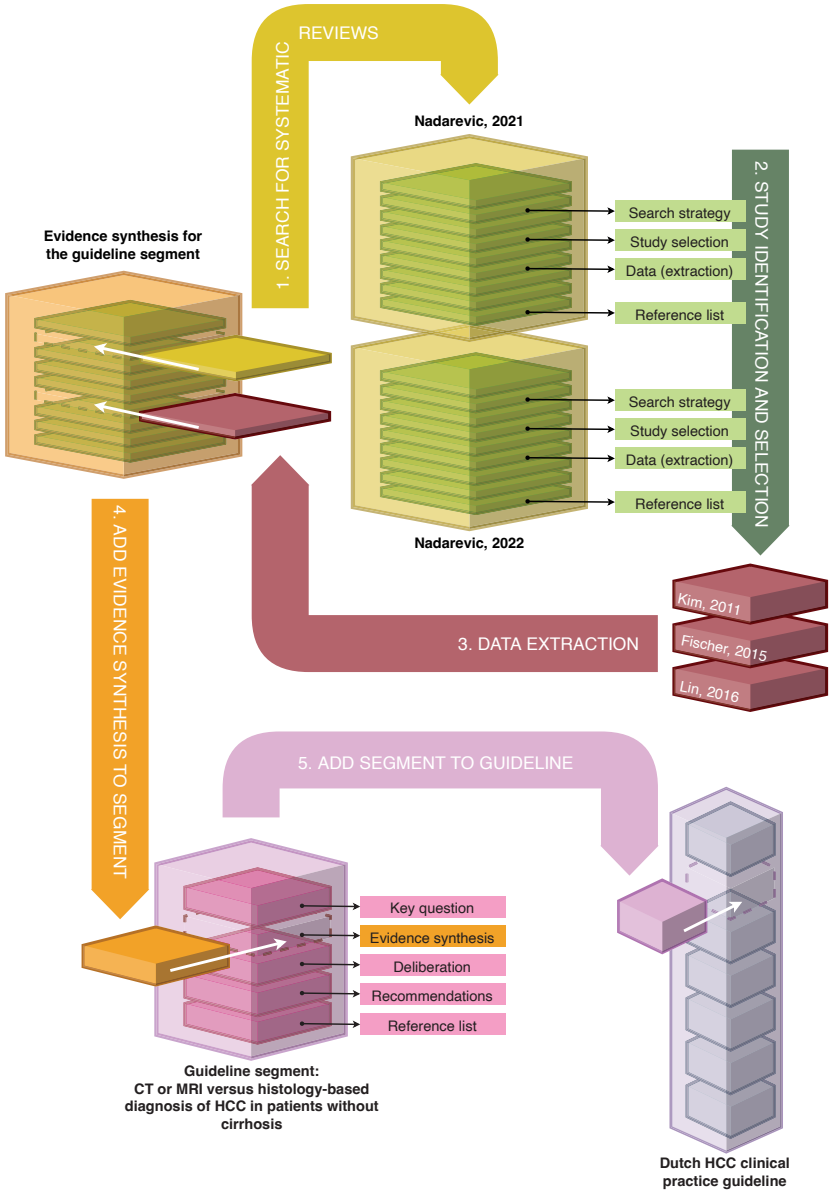


Figure 1. Visualization of the process leading to uptake of the studies in the Dutch HCC guideline. *We searched for published systematic reviews in our evidence synthesis for the Dutch hepatocellular carcinoma (HCC) clinical practice guideline segment (step 1). This resulted in the identification of two Cochrane diagnostic test accuracy reviews that also seemed to capture studies with a (sub)sample of suspected patients without liver cirrhosis. Through the transparently reported information and available data in these Cochrane reviews, we were able to identify three studies with direct evidence regarding the diagnostic test accuracy of computed tomography (CT) or magnetic resonance (MRI) in patients suspected of an HCC with a non-cirrhotic liver (step 2). Relevant data was extracted from these three studies for our own evidence synthesis (step 3). The evidence synthesis was then added to the guideline section for its development (step 4). Once endorsed, the guideline segment is added to the Dutch HCC clinical practice guideline (step 5) comprising of segments covering various other topics in the diagnosis and treatment of HCC (available via www.richtlijnendatabase.nl).*

of the PRISMA statement's items,[16] existing evidence indicates that adherence to these reporting guidelines is suboptimal.[17, 18] However, enhanced adherence does not only increase transparency and reproducibility of SRs but may, combined with sufficient data availability and accessibility, also increase the reusability of data within SRs. The integration of efforts between primary research, evidence synthesis, and clinical practice guideline development is crucial. Currently, these processes and efforts seem to be performed siloed.[19] In our approach, by being able to reuse efforts in the production of SRs and the data within these SRs, we have successfully bridged a notable gap between evidence synthesis and CPG development while optimizing resource utilization during the production of a CPG.

LIST OF ABBREVIATIONS

CPG: Clinical Practice Guideline
CT: Computed Tomography
HCC: Hepatocellular Carcinoma
MRI: Magnetic Resonance Imaging
SR: Systematic Review
TIAB: Title/Abstract

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The clinical practice guideline segment (in Dutch), including its evidence synthesis (in English), will be available through www.richtlijnendatabase.nl once the clinical practice guideline is endorsed in its finalization phase. The search string used in the search strategy will be available upon request via www.richtlijnendatabase.nl once the clinical practice guideline is endorsed and published. The Cochrane diagnostic test accuracy reviews and their statistical data files are stored in the Cochrane Database of Systematic Reviews (CDSR, ISSN 1469-493X) and can be found through the Cochrane Library at www.cochranelibrary.com.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

There was no specific funding received for this report.

AUTHORS' CONTRIBUTIONS

All author contributions are made transparent using the Contributor Role Taxonomy (CRediT), described at: https://doi.org/10.1002/leap.1210

MSO: conceptualization, methodology, investigation, visualization, writing – original draft
RdM: supervision (guideline), investigation, writing – review & editing
FGIvV: investigation, writing – review & editing
MWN: investigation, writing – review & editing
ET: investigation, writing – review & editing
CMWG: supervision (guideline), project administration (guideline), writing – review & editing
MJvdL: conceptualization, methodology, supervision (research), writing – review & editing.
LH: conceptualization, methodology, supervision (research), writing – review & editing

All authors were involved in reading and (re)writing the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Not applicable.

LITERATURE LIST

1. Netherlands Comprehensive Cancer Organisation. The Netherlands Cancer Registry data & figures application. 2023. Available from: https://iknl.nl/nkr-cijfers?lang%7Clanguage=en [Accessed on: 10 July 2023]
2. Statistics Netherlands. StatLine, Population dynamics; month and year. 2023. Available from: https://opendata.cbs.nl/statline/#/CBS/en/dataset/83474ENG/table?ts=1616407710127
3. Gawrieh S, Dakhoul L, Miller E, Scanga A, deLemos A, Kettler C, et al. Characteristics, aetiologies and trends of hepatocellular carcinoma in patients without cirrhosis: a United States multicentre study. Aliment Pharmacol Ther. 2019;50(7):809-21.
4. Yen YH, Cheng YF, Wang JH, Lin CC, Wang CC. Characteristics and etiologies of hepatocellular carcinoma in patients without cirrhosis: When East meets West. PLoS One. 2021;16(1):e0244939.
5. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol. 2018;69(1):182-236.
6. Singal AG, Llovet JM, Yarchoan M, Mehta N, Heimbach JK, Dawson LA, et al. AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology. 2023.
7. Nadarevic T, Colli A, Giljaca V, Fraquelli M, Casazza G, Manzotti C, et al. Magnetic resonance imaging for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. Cochrane Database Syst Rev. 2022;5(5):Cd014798.
8. Nadarevic T, Giljaca V, Colli A, Fraquelli M, Casazza G, Miletic D, Štimac D. Computed tomography for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. Cochrane Database Syst Rev. 2021;10(10):Cd013362.
9. Kim SE, Lee HC, Shim JH, Park HJ, Kim KM, Kim PN, et al. Noninvasive diagnostic criteria for hepatocellular carcinoma in hepatic masses >2 cm in a hepatitis B virus-endemic area. Liver Int. 2011;31(10):1468-76.
10. Lin MT, Wang CC, Cheng YF, Eng HL, Yen YH, Tsai MC, et al. Comprehensive Comparison of Multiple-Detector Computed Tomography and Dynamic Magnetic Resonance Imaging in the Diagnosis of Hepatocellular Carcinoma with Varying Degrees of Fibrosis. PLoS One. 2016;11(11):e0166157.
11. Fischer MA, Raptis DA, Donati OF, Hunziker R, Schade E, Sotiropoulos GC, et al. MR imaging features for improved diagnosis of hepatocellular carcinoma in the non-cirrhotic liver: Multi-center evaluation. Eur J Radiol. 2015;84(10):1879-87.
12. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022): Cochrane; 2023.
13. Vogel A, Cervantes A, Chau I, Daniele B, Llovet JM, Meyer T, et al. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(Suppl 4):iv238-iv55.
14. Moctezuma-Velázquez C, Lewis S, Lee K, Amodeo S, Llovet JM, Schwartz M, et al. Non-invasive imaging criteria for the diagnosis of hepatocellular carcinoma in non-

cirrhotic patients with chronic hepatitis B. JHEP Rep. 2021;3(6):100364.
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372:n71.
16. Tam WWS, Tang A, Woo B, Goh SYS. Perception of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement of authors publishing reviews in nursing journals: a cross-sectional online survey. BMJ Open. 2019;9(4):e026271.
17. Nawijn F, Ham WHW, Houwert RM, Groenwold RHH, Hietbrink F, Smeeing DPJ. Quality of reporting of systematic reviews and meta-analyses in emergency medicine based on the PRISMA statement. BMC Emerg Med. 2019;19(1):19.
18. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. Syst Rev. 2017;6(1):263.
19. 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. J Clin Epidemiol. 2020;123:153-61.

CHAPTER 8

Data sources and methods used to determine pretest probabilities in a cohort of Cochrane diagnostic test accuracy reviews

M.S. Oerbekke
K. Jenniskens
R.J.P.M. Scholten
L. Hooft

BMC Med Res Methodol. 2020 Apr 16;20(1):85
doi: 10.1186/s12874-020-00952-w
PMID: 32299367

ABSTRACT

BACKGROUND

A pretest probability must be selected to calculate data to help clinicians, guideline boards and policy makers interpret diagnostic accuracy parameters. When multiple analyses for the same target condition are compared, identical pretest probabilities might be selected to facilitate the comparison. Some pretest probabilities may lead to exaggerations of the patient harms or benefits, and guidance on how and why to select a specific pretest probability is minimally described. Therefore, the aim of this study was to assess the data sources and methods used in Cochrane diagnostic test accuracy (DTA) reviews for determining pretest probabilities to facilitate the interpretation of DTA parameters. A secondary aim was to assess the use of identical pretest probabilities to compare multiple meta-analyses within the same target condition.

METHODS

Cochrane DTA reviews presenting at least one meta-analytic estimate of the sensitivity and/or specificity as a primary analysis published between 2008 and January 2018 were included. Study selection and data extraction were performed by one author and checked by other authors. Observed data sources (e.g. studies in the review, or external sources) and methods to select pretest probabilities (e.g. median) were categorized.

RESULTS

Fifty-nine DTA reviews were included, comprising of 308 meta-analyses. A pretest probability was used in 148 analyses. Authors used included studies in the DTA review, external sources, and author consensus as data sources for the pretest probability. Measures of central tendency with or without a measure of dispersion were used to determine the pretest probabilities, with the median most commonly used. Thirty-two target conditions had at least one identical pretest probability for all of the meta-analyses within their target condition. About half of the used identical pretest probabilities were inside the prevalence ranges from all analyses within a target condition.

CONCLUSIONS

Multiple sources and methods were used to determine (identical) pretest probabilities in Cochrane DTA reviews. Indirectness and severity of downstream consequences may influence the acceptability of the certainty in calculated data with pretest probabilities. Consider: whether to present normalized frequencies, the influence of pretest probabilities on normalized frequencies, and whether to use identical pretest probabilities for meta-analyses in a target condition.

Keywords: Diagnostic test accuracy, pretest probability, pretest risk, background prevalence, review literature as topic, absolute numbers, normalized frequencies, natural frequencies

1. BACKGROUND

Diagnostic tests are essential to clinicians in their daily practice. Test results inform about the preferred healthcare pathway to, ideally, cure a patient from disease. The optimal way to understand a diagnostic test's performance and the downstream consequences for patients is through a test-treatment randomized controlled trial. Such trials provide comparative information on health outcomes (both harms and benefits) of healthcare pathways initiated by the outcome of the diagnostic tests or strategies. However, test-treatment randomized controlled trials are methodologically complex.[1] Diagnostic test accuracy (DTA) studies are usually an alternative to these complex trials and can be summarized in systematic reviews. DTA reviews include primary cross-sectional studies using the diagnostic test of interest and aggregate data by meta-analysis so that a pooled sensitivity and specificity is presented. Sensitivity is the proportion of persons with the target condition that are correctly classified by the index test, while specificity is the proportion of correctly classified persons without the target condition. Persons who are false negatively misclassified might not receive an intervention when they should have. Further diagnostic testing may be indicated which is possibly more burdensome or more harmful (e.g. the next diagnostic test is more invasive or may involve nuclear imaging). Persons who were false positively misclassified are falsely diagnosed with the presence of the target condition. Then, the provided intervention may be unnecessary and potentially burdensome, depending on the nature of the intervention. Complications from the intervention may arise and complaints may persist, potentially resulting in a late diagnosis of the actual present target condition. Further treatment for the actual target condition may then have more risks or complications compared to early diagnosis and intervention.

From literature it seems that clinicians have trouble interpreting accuracy parameters such as sensitivity and specificity.[2] To facilitate the interpretation of DTA results absolute numbers of true/false positives and true/false negatives can be presented in a hypothetical cohort of e.g. 1000 persons, which is also known as normalized frequencies.[2] However, to calculate normalized frequencies a pretest probability (i.e. the disease prevalence in the hypothetical cohort) needs to be determined. The normalized frequencies are then calculated and reported, whereafter the diagnostic test's end-user can interpret whether the test performance is acceptable in terms of true or false positives and negatives. Such normalized frequencies are usually presented in summary of findings tables in Cochrane DTA Reviews and in the evidence tables from the Grading of Recommendations Assessment, Development and Evaluation (GRADE).[3]

These normalized frequencies are not only important for clinicians, but also for guideline boards and policy makers. Decisions whether or not to recommend the use of a diagnostic test in a guideline or decisions about health care restitution may be influenced by the presented normalized frequencies. However, normalized frequencies are dependent on the chosen pretest probability. Using different pretest probabilities while the sensitivity and specificity remain constant may lead to an exaggeration of the estimated patient harms and/or benefits due to varying absolute numbers of (mis)classifications as calculated in the hypothetical cohort. Furthermore, the normalized frequencies from a hypothetical cohort can be more

directly compared when using identical pretest probabilities for multiple tests within a target condition. However, selecting an identical pretest probability for multiple meta-analyses in a single target condition could be challenging. The selected pretest probability might not lie in the prevalence range of every meta-analysis and, therefore, extrapolation of data may occur, potentially decreasing the certainty of the presented normalized frequencies.

While GRADE does not suggest a specific method to determine a pretest probability in its handbook[3, 4] the Cochrane Handbook does propose some methods (e.g. the median disease prevalence or the prevalence from disease registries) although a rationale to use a specific method is not given.[5] Because guidance in determining a pretest probability is minimally described, it is unknown what data sources (i.e. the data on which a pretest probability is based) and methods are actually used to determine a pretest probability in DTA reviews. Therefore, the aim of this study was to assess the data sources and methods used in a cohort of Cochrane DTA reviews to determine pretest probabilities for the facilitation of pooled DTA accuracy parameter interpretation and to provide some considerations for the use of normalized frequencies. A secondary aim was to assess the use of identical pretest probabilities in multiple analyses within the same target condition, necessary for the comparison of test performances.

2. METHODS

2.1 COHORT DEFINITION

The Cochrane Database of Systematic Reviews (CDSR) was accessed through the Cochrane Library. The Cochrane Collaboration has pioneered methods for DTA reviews and the CDSR contains DTA reviews since 2008.[6, 7] Cochrane DTA reviews published in the period from 2008 up to and including January 2018 were potentially eligible to enter the cohort. To obtain DTA reviews the CDSR was browsed by the topic 'Diagnosis', while protocols and intervention or methodology reviews were excluded through limiters in the search engine interface. A DTA review was included in the cohort when it reported at least one meta-analytic estimate of sensitivity and/or specificity (i.e. either retrieved with a bivariate model or by using a hierarchical summary receiver operating characteristic model) as a primary analysis in the presented tables, which were usually summary of findings tables. The screening and selection of eligible DTA reviews was performed by one author (MSO) and checked by the other authors (KJ, RJPMS, LH).

2.2 DATA EXTRACTION

The method for determining the pretest probability itself was recorded. For example, this could be the use of the mean or median disease prevalence (from studies included for a target condition). General characteristics (e.g. title, publication year), the number of meta-analyses in the review, whether a pretest probability was used, the number of pretest probabilities used (if applicable), the source of data for determining the pretest probability, and the method used for selecting pretest probabilities (if applicable) were extracted by one author (MSO)

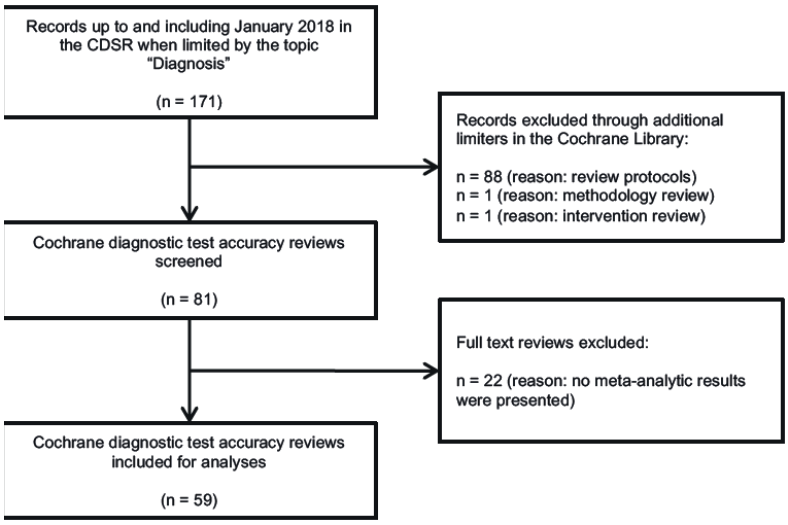


Figure 1. Cohort formation. *Flow diagram showing the formation of the cohort and reasons for exclusion. CDSR: Cochrane Database of Systematic Reviews.*

and checked by the other authors (KJ, RJPMS, LH). We only extracted data sources and methods for pretest probabilities from primary meta-analyses (usually presented in summary of findings tables). Meta-analyses with the purpose of being sensitivity analyses when excluding certain studies or with the purpose of heterogeneity analysis were not considered for data-extraction, even when they were presented in the summary of findings table. When a disease prevalence was reported but not used to interpret the sensitivity and/or specificity in some manner, the disease prevalence was not considered as a pretest probability. Descriptive statistics were performed in IBM SPSS Statistics for Windows (Version 21, 2012, Armonk, NY: IBM Corp.).

2.3 CATEGORIZATION OF DATA SOURCES

The first step in the categorization of extracted data was to divide the included meta-analyses into two groups; One group of analyses where no pretest probability was used and one group of analyses where a pretest probability was used. Next, for every meta-analysis that used a pretest probability the source of the pretest probability was determined. The pretest probability could be determined based on the studies that were included in the review, from external sources, or from author consensus. The data source of a pretest probability was categorized as unclear when the source could not be determined. Further categorization of data sources took place when the pretest probability was determined from included studies. Pretest probabilities used to interpret accuracy meta-analyses could then be determined from all studies included for the target condition, from all studies used per test/analysis for a target condition, from all included studies, from all included studies and from an unclear

source, from all included studies and from studies with a low risk of bias, or only from included studies that reported the disease prevalence. Further categorization also took place when the pretest probability was determined from external sources. Pretest probabilities could then be determined from published scientific literature, from the World Health Organization's suggestions, or from a guideline. The data source of pretest probabilities was categorized as 'author consensus' when a pretest probability was assumed by de review's authors and not based on included studies or external sources. See Additional File 2 for a description and example of each category. Methods to determine a pretest probability from the observed data sources were recorded and counted. Methods in external sources were not recorded, since these methods were not used in the Cochrane DTA reviews but in the external source.

2.4 IDENTICAL PRETEST PROBABILITIES WITHIN A TARGET CONDITION

Target conditions in DTA reviews that had multiple meta-analyses for the same target condition were identified to observe the use of identical pretest probabilities.

One author (MSO) extracted the following data from the multiple meta-analyses within a target condition: whether identical pretest probabilities were used for all of the meta-analyses within a target condition, and whether the prevalence ranges of all of the individual meta-analyses for a target condition contained the selected identical pretest probability or not. When the prevalence range of individual meta-analyses in a single target condition did not contain the selected identical pretest probability, a justification by the review authors was sought in the review. Data extraction was checked by a second author (KJ). Prevalence ranges in individual meta-analyses were calculated from data in the DTA review's appendices when not reported in the text.

3. RESULTS

3.1 COHORT DESCRIPTION

The CDSR contained 171 documents on the topic 'Diagnosis'. There were 81 reviews left after excluding 88 review protocols and 2 reviews on interventions and methodology. After screening the full text an additional 22 DTA reviews were excluded as no meta-analytic results were presented (referenced in Additional File 1). Consequently, 59 Cochrane DTA reviews were included in the cohort (Figure 1 and Additional File 1). The 59 DTA reviews in the cohort contained 308 meta-analyses (see Table 1). The number of meta-analyses ranged from 1 to 34 (median: 3) per review. In 16 reviews (16/59, 27.1%) there were 150 meta-analyses (150/308, 48.7%) that did not use a pretest probability. Thirty-nine reviews (39/59, 66.1%) had 143 meta-analyses (143/308, 46.4%) where a pretest probability was used. Four reviews (4/59, 6.8%) contained 15 meta-analyses for which a pretest probability was used in 5 analyses. Therefore, a total of 160 analyses (160/308, 51.9%) were found where no pretest probability was used and 148 analyses (148/308, 48.1%) were found where at least one pretest probability was used.

Table 1 – General characteristics

	Reviews (n)	Meta-analyses (n [% using a pretest probability])	Number of meta-analyses per DTA review (median [range])	Number of pretest probabilities used per meta-analysis[a] (median [range])
Cochrane DTA reviews				
Total included DTA reviews	59	308 (48.1)	3 (1-34)	1 (1-6)
Reviews not using a pretest probability at all	16	150 (0)	4 (2-34)	-[a]
Reviews using a pretest probability for all pooled analyses	39	143 (100)	3 (1-16)	1 (1-5)
Reviews reporting analyses with and without pretest probabilities	4	15 (33.3)	3.5 (3-5)	3 (1-6)

DTA: Diagnostic Test Accuracy, IQR: Interquartile Range
[a]Could not be calculated since there were no analyses using a pretest probability

3.2 SOURCES OF PRETEST PROBABILITIES

In the 148 analyses in which a pretest probability was used three main categories of data sources were distinguished (Figure 2). In 90 (60.8%) of the 148 analyses the pretest probability was determined from included studies, in 26 analyses (26/148, 17.6%) from external sources, and from author consensus in 31 analyses (31/148, 20.9%). In one analysis the data source was unclear. When the included studies in the review were used to determine one or multiple pretest probabilities, the data source could be further differentiated. A pretest probability was determined from all studies included for a target condition in 40 analyses (40/90, 44.4%), from studies used per test/analysis for a target condition in 22 analyses (22/90, 24.4%), or from all studies in the systematic review in 18 analyses (18/90, 20%). Five other analyses (5/90, 5.6%) had multiple pretest probabilities where some were determined from all studies in the systematic reviews and some from an unclear source. Three analyses (3/90, 3.3%) had multiple pretest probabilities, where some were obtained from all studies in the systematic review and some from the included low risk of bias studies. A pretest probability was determined in 2 analyses (2/90, 2.2%) from the studies that reported their prevalence. When the pretest probability was obtained from external sources, the data sources could also be further differentiated. In 14 analyses (14/26, 53.8%) the pretest probability was a disease prevalence reported in published scientific literature, in 10 analyses (10/26, 38.5%) a suggestion by the WHO, and in 2 analyses (2/26, 7.7%) a disease prevalence reported in a guideline.

3.3 METHODS FOR DETERMINING A PRETEST PROBABILITY

Pretest probabilities based on studies within the review were determined by using measures of central tendency (e.g. median) whether or not combined with measures of dispersion (e.g.

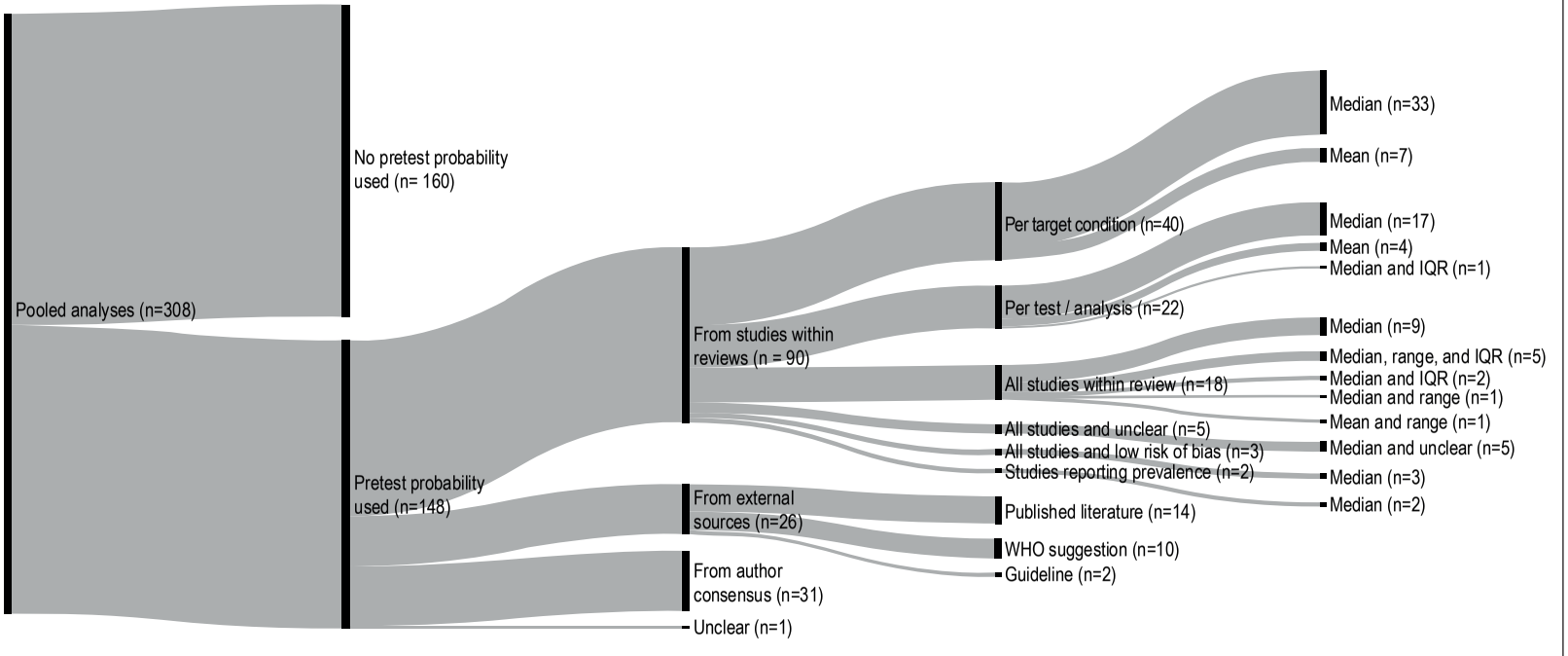


Figure 2. Data sources and methods for determining pretest probabilities in Cochrane Diagnostic Test Accuracy reviews. *Sankey plot showing which data sources and methods were used to determine the pretest probability in Cochrane Diagnostic Test Accuracy reviews. IQR: Interquartile Range, WHO: World Health Organization*

range). Using multiple methods resulted in multiple pretest probabilities for a single analysis (e.g. using a median with a range results in three pretest probabilities). The median was used individually in 64 analyses (64/90, 71.1%) or together with the interquartile range in 3 analyses (3/90, 3.3%), with the range in 1 analysis (1/90, 1.1%), with the range and interquartile range in 5 analyses (5/90, 5.6%), or together with other estimates of pretest probabilities for which the method was unclear in 5 analyses (5/90, 5.6%). The mean was used individually in 7 analyses (7/90, 7.8%) or together with the range in 1 analysis (1/90, 1.1%). Figure 2 shows the methods per data source and the number of analyses in where these methods were used.

3.4 IDENTICAL PRETEST PROBABILITIES FOR MULTIPLE META-ANALYSES WITHIN TARGET CONDITIONS

There were 41 target conditions having two or more analyses which might have used identical pretest probabilities to compare test performances for tests in the same diagnostic role within a target condition (Figure 3). Nine target conditions (9/41, 22%) did not have identical pretest probabilities in their analyses. One target condition (1/41, 2.4%) had six identical pretest probabilities for its two meta-analyses, however it was unclear whether the prevalence ranges of the two meta-analyses contained the pretest probability estimates. Thirty-two target conditions (32/41, 78%) had at least one identical pretest probability (range: 1-6) that was used for all of the meta-analyses within their respective target condition. Sixty-nine identical pretest probabilities within target conditions were identified. Thirty-seven pretest probabilities (37/69, 53.6%) fell in the prevalence range of all of the individual meta-analyses within that target condition. However, 26 pretest probabilities (26/69, 37.7%) in 11 DTA reviews fell outside the prevalence range from at least one meta-analysis within that target condition. In 5 of these 11 DTA reviews pretest probabilities were used to show test performances in different scenarios (i.e. referral scenarios, low/high risk scenarios, newly/previously treated cases, geographical locations, adults/children).

4. DISCUSSION

4.1 SUMMARY OF KEY FINDINGS

A total of 59 Cochrane DTA reviews were included to assess the data sources and methods used to determine pretest probabilities and to assess the use of identical pretest probabilities in multiple analyses within a target condition. Various sources and methods to determine a pretest probability were found. Sixteen DTA reviews did not use a single pretest probability. Almost half of the observed meta-analyses used at least one pretest probability (range: 1-6 pretest probabilities) to facilitate the interpretability of the results. The median was the most used method to determine a pretest probability. Thirty-nine target conditions contained two or more analyses and used at least one identical pretest probability for all of its analyses (range: 1-6 pretest probabilities). Twenty-six of the identical pretest probabilities (37.7%) fell outside the disease prevalence range of at least one analysis within the target condition.

4.2 INTERPRETATION OF RESULTS

The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy proposes to use the median prevalence of the included studies or external sources as methods to select a pretest probability.[5] Indeed, the median was used more than any other method in Cochrane DTA reviews when the pretest probability was determined based on studied included in the review. External sources were also used in 26 analyses. The Cochrane Handbook also suggests the use of disease registries for selecting a representative pretest probability, but no such method was specifically mentioned in the included Cochrane DTA reviews. Observed prevalence from disease surveillance systems and epidemiological surveys by the WHO, however, might be interpreted as disease registries as well. The Cochrane Handbook states that a representative pretest probability may be derived from the included studies only when the studies are representative for the target setting,[5] in which case the selected pretest probability will fall within the prevalence range of included studies. However, a pretest probability may be selected that falls outside the disease prevalence range from the included studies. This might also be the case when authors wish to use an identical pretest probability for several meta-analyses within the same target condition and use, for example, the median prevalence from all included studies for that target condition (e.g. the median of all included studies may fall outside some prevalence ranges when these ranges do not overlap sufficiently). A pretest probability could be representative for the target setting, but it might not necessarily be an appropriate pretest probability in the context of the disease prevalence range from the included studies in a meta-analysis. When a representative pretest probability falls outside the disease prevalence range it might be considered indirect, since the observed data did not contain that pretest probability. When considering the observed data, a more data-appropriate pretest probability from within the disease prevalence range from the meta-analysis itself might be selected. Therefore, a representative and a data-appropriate pretest probability are not necessarily the same (see Example 1 in Additional File 3). Appropriateness for the data in this case means to not extrapolate outside of the data in the meta-analysis, because there is uncertainty about the test's performance outside the observed data.

4.3 STUDY LIMITATIONS

A potential limitation of this study is that only Cochrane DTA reviews were included. Since the Cochrane Handbook proposes to use the median it was beforehand likely to observe that the median is being preferred in Cochrane DTA reviews. Different methods and data sources for determining pretest probabilities in non-Cochrane DTA reviews could have been missed. Furthermore, DTA review authors may also choose to aid the interpretation of non-meta-analytic results. For example, normalized frequencies may also have been calculated for single DTA studies in the review. However, only the use of pretest probabilities in meta-analyses was assessed in the current study. The potential issues regarding the data sources and methods for choosing a pretest probability remain the same for single studies as for meta-analytic estimates. Data sources and methods for determining pretest probabilities in the absence of a meta-analysis are therefore still unknown. However, when there were data sources or methods that this study did not address, it adds to the impression that there is no consensus on what data source and method to use for determining a representative or

appropriate pretest probability.

4.4 IMPLICATIONS FOR USING PRETEST PROBABILITIES

From the results of this study no clear guidance can be given on what source or what method should be used for determining a pretest probability. Furthermore, it is unknown if a pretest probability outside the disease prevalence range is problematic in clinical reality. Whether or not it is problematic may also be context dependent, as for some target conditions a certain number of misclassifications are more acceptable than for target conditions where misclassifications have severe downstream consequences. Even if it turns out to be clinically problematic in future research, it presently might still be best practice to facilitate the interpretability of diagnostic accuracy parameters by presenting normalized frequencies. Furthermore, there are some considerations which may be taken in to account when presenting results in DTA reviews.

First to consider is whether to provide a way for end-users to interpret the presented accuracy parameters, as it was observed in this study that about half of all meta-analyses were not accompanied with normalized frequencies from a hypothetical cohort. Literature shows that interpreting diagnostic test accuracy parameters may be troublesome for its users and therefore normalized frequencies may be useful.[2] However, choosing pretest probabilities to calculate normalized frequencies is not without difficulties and therefore it is uncertain whether normalized frequencies are trustworthy enough for all decision-making (see the second consideration). The need for interpretability versus the certainty of and need for a truthful representation might determine whether normalized frequencies are calculated. We might accept more or less certainty in the normalized frequencies depending on their degree of indirectness and the severity of downstream consequences for misclassifications (see Example 2 in Additional File 3). However, not facilitating the interpretation of accuracy parameters also complicates the judgement by clinicians, policy makers, and guideline boards whether to use the diagnostic test.

Secondly, consider giving thought about the influence of the method of selecting the pretest probability on the normalized frequencies from the hypothetical cohort. A guideline board may base their decision about whether or not to recommend a test for clinical practice on the presented normalized frequencies. It is important to understand that different pretest probabilities will result in different normalized frequencies while the sensitivity and specificity remain constant (see Example 3 in Additional File 3), potentially influencing the decision-making in practice, policy or guidelines.

Thirdly, when there are multiple meta-analyses for the same target condition, consider whether to use an identical pretest probability in each of those analyses so that the normalized frequencies can be compared. Ideally the selected pretest probabilities fall inside all of the disease prevalence ranges from all individual meta-analyses within the target condition, although this might not be feasible for every scenario (e.g. when the disease prevalence ranges from the meta-analyses do not overlap). However, from this study no guidance can be provided on whether an identical pretest probability is suitable for all of the disease prevalence ranges in the analyses, even when the pretest probability falls inside all prevalence ranges.

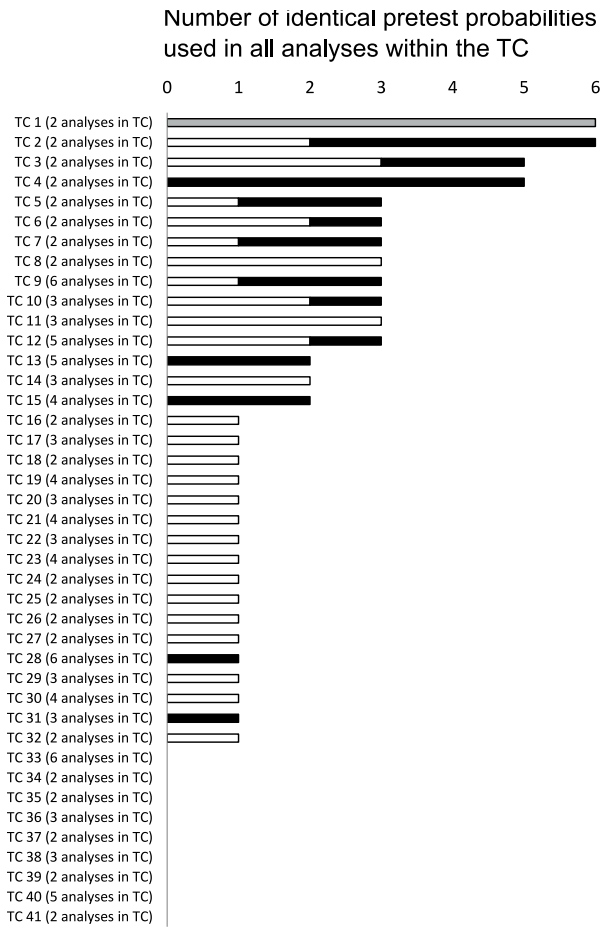


Figure 3. Identical pretest probabilities within target conditions. *The figure shows the number of target conditions with two or more analyses and the use of identical pretest probabilities for all of the meta-analyses within a target condition. Black bars indicate the number of pretest probabilities that were outside the prevalence range of at least one meta-analysis for the target condition. White bars indicate the number of pretest probabilities that were inside the prevalence ranges of all individual meta-analyses for that specific target condition. The gray bar indicates that it was unclear whether these identical pretest probabilities were inside or outside the disease prevalence ranges. No bars were drawn in the figure when there were no identical pretest probabilities used for all of the meta-analyses within a target condition. DTA: Diagnostic Test Accuracy, TC: Target Condition*

Providing clinicians, policy makers, and guideline boards with methods to facilitate the interpretation of DTA results is not only important for them, but ultimately also for patients who undergo diagnostic tests. Different pretest probabilities will result in different normalized frequencies. However, it is not known whether differences in normalized frequencies caused by the use of different pretest probabilities actually impacts decision-making and whether it will then clinically harm of benefit patients. The future direction of research in this area could focus on whether different pretest probabilities will actually result in a different clinical decision, guideline recommendation, or policy change. Furthermore, future research could focus on developing other strategies for accuracy parameters so that they are both interpretable and helpful when research shows that calculating normalized frequencies may not be beneficial for actual decision-making by clinicians, policy makers, or guideline boards.

5. CONCLUSIONS

Various data sources and methods are used to obtain pretest probabilities, without consensus on which data source or method to use. Identical pretest probabilities were used in some DTA reviews, where test performances for a target condition could be directly compared in about half of the identical pretest probabilities. The certainty of presented data calculated with pretest probabilities is influenced by indirectness. Indirectness is probably more acceptable in situations where there are less severe downstream consequences, but the reduction of certainty should be acknowledged. There are three considerations that might be taken in to account when presenting DTA results: consider whether or not to present normalized frequencies from a hypothetical cohort; consider the influence of the chosen method for selecting a pretest probability on the normalized frequencies on which a clinical decision, guideline recommendation or policy change may be based, and consider to use identical pretest probabilities that fall within the range of the selected studies when there are multiple meta-analyses for the same target condition.

LIST OF ABBREVIATIONS

CDSR: Cochrane Database of Systematic Reviews
DTA: Diagnostic Test Accuracy
GRADE: Grading of Recommendations, Assessment, Development and Evaluation
WHO: World Health Organization

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethics approval was not required for the conduct of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIAL

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

COMPETING INTERESTS

Authors declare that they have no competing interests.

FUNDING

No funding was received for the conduct of this research.

AUTHORS' CONTRIBUTION

All authors contributed equally to the intellectual development of the manuscript. The first author (MS Oerbekke) did the data-extraction which was checked by the other authors (K Jenniskens, RJPM Scholten, L Hooft). Data-analyses and the construction of tables and figures were performed by the first author (MS Oerbekke). Drafts of the manuscript were provided by the first author (M Oerbekke) and were co-developed by the other authors (K Jenniskens, RJPM Scholten, L Hooft). All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

We would like to thank dr. W.A. Van Enst for her efforts and contributions regarding data-checking.

LITERATURE LIST

1. Ferrante di Ruffano L, Dinnes J, Sitch AJ, Hyde C, Deeks JJ. Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests. BMC medical research methodology. 2017;17(1):35.
2. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JA, Hyde C, Ben-Shlomo Y. How well do health professionals interpret diagnostic information? A systematic review. BMJ open. 2015;5(7):e008155.
3. Mustafa RA, Wiercioch W, Santesso N, Cheung A, Prediger B, Baldeh T, et al. Decision-Making about Healthcare Related Tests and Diagnostic Strategies: User Testing of GRADE Evidence Tables. PloS one. 2015;10(10):e0134553.
4. Schüneman H, Brożek J, Guyatt G, Oxman A, editors. Chapter 7: The GRADE Approach for Diagnostic Tests and Strategies. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach, Updated October 2013Updated October 2013. Available from https://guidelinedevelopment.org/handbook.
5. Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11: Interpreting results and drawing conclusions. In: Deeks J, Bossuyt P, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagostic Test Accuracy, Version 092013, Available from: http://methods.cochrane.org/sdt/handbook-dta-reviews.
6. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008;149(12):889-97.
7. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. Syst Rev. 2013;2:82.

CHAPTER 8
Additional files

ADDITIONAL FILE 1

INCLUDED AND EXCLUDED REVIEWS

Author	Year	Name	url	Reason for exclusion
EXCLUDED REVIEWS				
Hull et al.	2017	Tests for detecting strabismus in children aged 1 to 6 years in the community	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011221.pub2/full	No pooled/summary analysis performed
Davidson et al.	2017	Amylase in drain fluid for the diagnosis of pancreatic leak in post-pancreatic resection	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012009.pub2/full	No pooled/summary analysis performed
Mens et al.	2017	Imaging for the exclusion of pulmonary embolism in pregnancy	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011053.pub2/full	No pooled/summary analysis performed
Harrison et al.	2016	Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011333.pub2/full	No pooled/summary analysis performed
Crawford et al.	2016	Ankle brachial index for the diagnosis of lower limb peripheral arterial disease	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010680.pub2/full	No pooled/summary analysis performed
Crawford et al.	2016	D-dimer test for excluding the diagnosis of pulmonary embolism	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010864.pub2/full	No pooled/summary analysis performed
Nisenblat et al.	2016	Combination of the non-invasive tests for the diagnosis of endometriosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012281/full	No pooled/summary analysis performed
Liu et al.	2015	Urinary biomarkers for the non-invasive diagnosis of endometriosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012019/full	No pooled/summary analysis performed
Davis et al.	2015	Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010775.pub2/full	No pooled/summary analysis performed
Bleeker et al.	2015	123I-MIBG scintigraphy and 18F-FDG-PET imaging for diagnosing neuroblastoma	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009263.pub2/full	No pooled/summary analysis performed

Palaniyappan et al.	2015	Voxel-based morphometry for separation of schizophrenia from other types of psychosis in first episode psychosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011021.pub2/full	No pooled/summary analysis performed
Archer et al.	2015	Regional Cerebral Blood Flow Single Photon Emission Computed Tomography for detection of Frontotemporal dementia in people with suspected dementia	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010896.pub2/full	No pooled/summary analysis performed
Arevalo et al.	2015	Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010783.pub2/full	No pooled/summary analysis performed
Hunt et al.	2015	Thromboelastography (TEG) and rotational thromboelastometry (ROTEM) for traumainduced coagulopathy in adult trauma patients with bleeding	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010438.pub2/full	No pooled/summary analysis performed
Fage et al.	2015	Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010860.pub2/full	No pooled/summary analysis performed
McCleery et al.	2015	Dopamine transporter imaging for the diagnosis of dementia with Lewy bodies	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010633.pub2/full	No pooled/summary analysis performed
Harrison et al.	2014	Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010771.pub2/full	No pooled/summary analysis performed
Rutten et al.	2014	Laparoscopy for diagnosing resectability of disease in patients with advanced ovarian cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009786.pub2/full	No pooled/summary analysis performed
Walsh et al.	2013	Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010173.pub2/full	No pooled/summary analysis performed
Hanchard et al.	2013	Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007427.pub2/full	No pooled/summary analysis performed
Henschke et al.	2013	Red flags to screen for malignancy in patients with low-back pain	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008686.pub2/full	No pooled/summary analysis performed

Wiliams et al.	2013	Red flags to screen for vertebral fracture in patients presenting with low-back pain	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008643.pub2/full	No pooled/summary analysis performed
INCLUDED REVIEWS				
Koliopoulos et al.	2017	Cytology versus HPV testing for cervical cancer screening in the general population	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008587.pub2/full	
Abraha et al.	2017	Ultrasonography for endoleak detection after endoluminal abdominal aortic aneurysm repair	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010296.pub2/full	
Wijedoru et al.	2017	Rapid diagnostic tests for typhoid and paratyphoid (enteric) fever	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008892.pub2/full	
Nieuwenhuis et al.	2017	Three-dimensional saline infusion sonography compared to two-dimensional saline infusion sonography for the diagnosis of focal intracavitary lesions	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011126.pub2/full	
Colli et al.	2017	Platelet count, spleen length, and platelet count-to-spleen length ratio for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008759.pub2/full	
Rompianesi et al.	2017	Serum amylase and lipase and urinary trypsinogen and amylase for diagnosis of acute pancreatitis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012010.pub2/full	
Best et al.	2017	Imaging modalities for characterising focal pancreatic lesions	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010213.pub2/full	
Ritchie et al.	2017	CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010803.pub2/full	
Pammi et al.	2017	Molecular assays for the diagnosis of sepsis in neonates	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011926.pub2/full	
Tamburrino et al.	2016	Diagnostic accuracy of different imaging modalities following computed tomography (CT) scanning for assessing the resectability with curative intent in pancreatic and periampullary cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011515.pub2/full	

Theron et al.	2016	GenoType® MTBDRsl assay for resistance to second-line anti-tuberculosis drugs	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010705.pub3/full
Allen et al.	2016	Diagnostic accuracy of laparoscopy following computed tomography (CT) scanning for assessing the resectability with curative intent in pancreatic and periampullary cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009323.pub3/full
Shaikh et al.	2016	Dimercaptosuccinic acid scan or ultrasound in screening for vesicoureteral reflux among children with urinary tract infections	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010657.pub2/full
Cohen et al.	2016	Rapid antigen detection test for group A streptococcus in children with pharyngitis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010502.pub2/full
Shah et al.	2016	Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011420.pub2/full
Nisenblat et al.	2016	Blood biomarkers for the non-invasive diagnosis of endometriosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012179/full
Gupta et al.	2016	Endometrial biomarkers for the non-invasive diagnosis of endometriosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012165/full
Ratnavelu et al.	2016	Intraoperative frozen section analysis for the diagnosis of early stage ovarian cancer in suspicious pelvic masses	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010360.pub2/full
Nisenblat et al.	2016	Imaging modalities for the non-invasive diagnosis of endometriosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009591.pub2/full
Creavin et al.	2016	Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011145.pub2/full
Leeflang et al.	2015	Galactomannan detection for invasive aspergillosis in immunocompromised patients	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007394.pub2/full
Nicholson et al.	2015	Blood CEA levels for detecting recurrent colorectal cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011134.pub2/full
Alldred et al.	2015	Urine tests for Down's syndrome screening	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011984/full

Alldred et al.	2015	First trimester serum tests for Down's syndrome screening	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011975/full
Michelessi et al.	2015	Optic nerve head and fibre layer imaging for diagnosing glaucoma	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008803.pub2/full
Cruciani et al.	2015	Polymerase chain reaction blood tests for the diagnosis of invasive aspergillosis in immunocompromised people	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009551.pub3/full
Mallee et al.	2015	Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010023.pub2/full
Macey et al.	2015	Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010276.pub2/full
Hooper et al.	2015	Clinical symptoms, signs and tests for identification of impending and current water-loss dehydration in older people	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009647.pub2/full
Ochodo et al.	2015	Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009579.pub2/full
Harrison et al.	2015	Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010772.pub2/full
Gurusamy et al.	2015	Endoscopic retrograde cholangiopancreatography versus intraoperative cholangiography for diagnosis of common bile duct stones	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010339.pub2/full
Gurusamy et al.	2015	Ultrasound versus liver function tests for diagnosis of common bile duct stones	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011548/full
Giljaca et al.	2015	Endoscopic ultrasound versus magnetic resonance cholangiopancreatography for common bile duct stones	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011549/full

Mocellin et al.	2015	Diagnostic accuracy of endoscopic ultrasonography (EUS) for the preoperative locoregional staging of primary gastric cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009944.pub2/full
Smailagic et al.	2015	18F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010632.pub2/full
Soares et al.	2015	First rank symptoms for schizophrenia	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010653.pub2/full
Pavlov et al.	2015	Transient elastography for diagnosis of stages of hepatic fibrosis and cirrhosis in people with alcoholic liver disease	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010542.pub2/full
Shaikh et al.	2015	Procalcitonin, C-reactive protein, and erythrocyte sedimentation rate for the diagnosis of acute pyelonephritis in children	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009185.pub2/full
Virgili et al.	2015	Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008081.pub3/full
Abba et al.	2014	Rapid diagnostic tests for diagnosing uncomplicated non-falciparum or Plasmodium vivax malaria in endemic countries	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011431/full
Schmidt et al.	2014	PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009519.pub2/full
Colli et al.	2014	Capsule endoscopy for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008760.pub2/full
Josephson et al.	2014	Computed tomography angiography or magnetic resonance angiography for detection of intracranial vascular malformations in patients with intracerebral haemorrhage	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009372.pub2/full
Zhang et al.	2014	11C-PIB-PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010386.pub2/full

Lawrie et al.	2014	Sentinel node assessment for diagnosis of groin lymph node involvement in vulval cancer	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010409.pub2/full
Boelaert et al.	2014	Rapid tests for the diagnosis of visceral leishmaniasis in patients with suspected disease	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009135.pub2/full
Ritchie et al.	2014	Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008782.pub4/full
Quinn et al.	2014	Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010079.pub2/full
Taylor et al.	2014	Computed tomography (CT) angiography for confirmation of the clinical diagnosis of brain death	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009694.pub2/full
Steingart et al.	2014	Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009593.pub3/full
Lenza et al.	2013	Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009020.pub2/full
Arbyn et al.	2013	Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008054.pub2/full
Wang et al.	2012	Clinical symptoms and signs for the diagnosis of Mycoplasma pneumoniae in children and adolescents with community-acquired pneumonia	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009175.pub2/full
Alldred et al.	2012	Second trimester serum tests for Down's Syndrome screening	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009925/full
Wang et al.	2011	Cardiac testing for coronary artery disease in potential kidney transplant recipients	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008691.pub2/full

Abba et al.	2011	Rapid diagnostic tests for diagnosing uncomplicated P. falciparum malaria in endemic countries	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008122.pub2/full
Windt et al.	2010	Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007431.pub2/full
Brazzelli et al.	2009	Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms	http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007424.pub2/full

ADDITIONAL FILE 2

EXAMPLES OF DATA SOURCE CATEGORIES

FROM INCLUDED STUDIES

From all studies included for the target condition

There are multiple target conditions in the systematic review and a target condition has one or multiple index tests. All studies from all analyses for a single target condition were used to determine a pretest probability. See Lenza et al. (2013) for an example.1

From studies used per test/analysis for a target condition

A target condition has multiple index tests. For each analysis of the index test the included studies for that index test were used to determine a pretest probability. See Colli et al. (2017) for an example.[2]

From all studies in the systematic review across all target conditions

A review used all of the included studies for analyses across all of the target conditions to determine a pretest probability. Analyses were also placed in this category if there was only one target condition defined in the systematic review and a pretest probability was determined from all included studies for that target condition, unless specifically stated otherwise by the authors. See Leeflang et al. (2015) for an example.[3]

From all studies in the systematic review and from an unclear source

An analysis used two pretest probabilities to calculate normalized frequencies. One of the pretest probabilities was determined from all of the included studies for analyses across all target conditions, while the other pretest probability had an unclear data source. See Abba et al. (2011) for an example.[4]

From all studies in the systematic review and only from studies with a low risk of bias

An analysis used two pretest probabilities to calculate normalized frequencies. One pretest probability was determined by all of the included studies for analyses across all target conditions, while the other pretest probability was calculated solely from studies with a low risk of bias. See Colli at al. (2014) for an example.[5]

Only from included studies that reported the disease prevalence

Only studies that reported their sample's disease prevalence were used to determine the pretest probability. See Ritchie et al. (2017) for an example.[6]

FROM EXTERNAL SOURCES

From published scientific literature

The pretest probability used in the systematic review was based on or informed by the disease prevalence as reported in published scientific literature. See Wijedoru et al. (2017) for an example.[7]

From a WHO suggestion

The pretest probability used in the systematic review was based on or informed by the

disease prevalence as suggested by the WHO. See Steingart et al. (2014) for an example.[8]

From a guideline

The pretest probability used in the systematic review was based on or informed by the disease prevalence as reported in a guideline. See Wang et al. (2011) for an example.[9]

AUTHOR CONSENSUS

It was considered author consensus when authors determined a pretest probability based on an assumption while not calculated from included studies or taken from external sources (including suggestions from external parties). See Shaikj et al. (2016) for an example.[10]

LITERATURE LIST

1. Lenza M, Buchbinder R, Takwoingi Y, Johnston RV, Hanchard NC, Faloppa F. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. The Cochrane database of systematic reviews. 2013(9):Cd009020.
2. Colli A, Gana JC, Yap J, Adams-Webber T, Rashkovan N, Ling SC, Casazza G. Platelet count, spleen length, and platelet count-to-spleen length ratio for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis. The Cochrane database of systematic reviews. 2017;4:Cd008759.
3. Leeflang MM, Debets-Ossenkopp YJ, Wang J, Visser CE, Scholten RJ, Hooft L, et al. Galactomannan detection for invasive aspergillosis in immunocompromised patients. The Cochrane database of systematic reviews. 2015(12):Cd007394.
4. Abba K, Deeks JJ, Olliaro P, Naing CM, Jackson SM, Takwoingi Y, et al. Rapid diagnostic tests for diagnosing uncomplicated P. falciparum malaria in endemic countries. The Cochrane database of systematic reviews. 2011(7):Cd008122.
5. Colli A, Gana JC, Turner D, Yap J, Adams-Webber T, Ling SC, Casazza G. Capsule endoscopy for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis. The Cochrane database of systematic reviews. 2014(10):Cd008760.
6. Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). The Cochrane database of systematic reviews. 2017;3:Cd010803.
7. Wijedoru L, Mallett S, Parry CM. Rapid diagnostic tests for typhoid and paratyphoid (enteric) fever. The Cochrane database of systematic reviews. 2017;5:Cd008892.
8. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. The Cochrane database of systematic reviews. 2014(1):Cd009593.
9. Wang LW, Fahim MA, Hayen A, Mitchell RL, Baines L, Lord S, et al. Cardiac testing for coronary artery disease in potential kidney transplant recipients. The Cochrane database of systematic reviews. 2011(12):Cd008691.
10. Shaikh N, Spingarn RB, Hum SW. Dimercaptosuccinic acid scan or ultrasound in screening for vesicoureteral reflux among children with urinary tract infections. The Cochrane database of systematic reviews. 2016;7:Cd010657.

ADDITIONAL FILE 3

EXAMPLES

EXAMPLE 1 – REPRESENTATIVENESS FOR A SETTING AND DATA-APPROPRIATENESS

DTA review authors may choose to use a pretest probability that is representative for the setting in which the diagnostic test is used. However, this representative pretest probability may fall outside the disease prevalence range from the observed data. External estimates of prevalence might be used when included studies do not match the population to which the test is to be applied in practice. The review authors then need to justify their choice and should acknowledge that there might be more uncertainty in their conclusion in relation to the downstream consequences of using the test in a hypothetical cohort. Although the pretest probability is representative for their target population (or setting), the meta-analysis of test performance measures includes studies that potentially do not closely match their situation. We know that patient spectrum affects prevalence, and therefore also performance measures as sensitivity and specificity. Using external pretest probabilities may lead to exaggeration or underestimation of the patient harms or benefits in the hypothetical cohort. Even when there is high confidence in the pooled accuracy estimates (e.g. assessed with GRADE), there might possibly be some indirectness in the calculations made with pretest probabilities outside the disease prevalence range.

For example, in the Cochrane DTA review by Wijedoru et al. (2017)[1] there were 5 meta-analytic results presented in the summary of findings table for diagnostic tests in (para) typhoid fever. Three representative pretest probabilities were selected by the review authors for geographical regions and populations: 1% for children in Africa, 10% for adults and children in Africa, and 30% for Asia. The prevalence range for each meta-analysis is shown in Figure A1. From the figure we can see that the prevalence ranges of all meta-analyses contain the representative pretest probabilities of 10% and 30%. Both pretest probabilities are appropriate for the observed data as well, since the prevalence ranges of all the meta-analyses contain both 10% and 30%. For children in Africa the representative pretest probability is 1%. The prevalence ranges from three meta-analyses contain this representative pretest probability and are, therefore, appropriate for the data in these three analyses. The representative pretest probability of 1% is not appropriate in the context of the observed data in the Typhidot (subset 2) and Test-it analyses, since the disease prevalence ranges of both meta-analyses did not contain the representative pretest probability of 1%. Therefore, the normalized frequencies are calculated using the pooled accuracy estimates that are not based on data about the test's performance at a prevalence of 1%. Nonetheless, the choice and justification to use the pretest probability of 1% is legitimate and it is unsure whether presenting normalized frequencies with pretest probabilities outside the prevalence range is actually harmful in clinical reality.

EXAMPLE 2 – THE NEED FOR INTERPRETABILITY VERSUS THE CERTAINTY OF THE DATA

Several situations might lower the confidence in the certainty of calculated downstream consequences (effects) of test usage. However, the need for high certainty/confidence

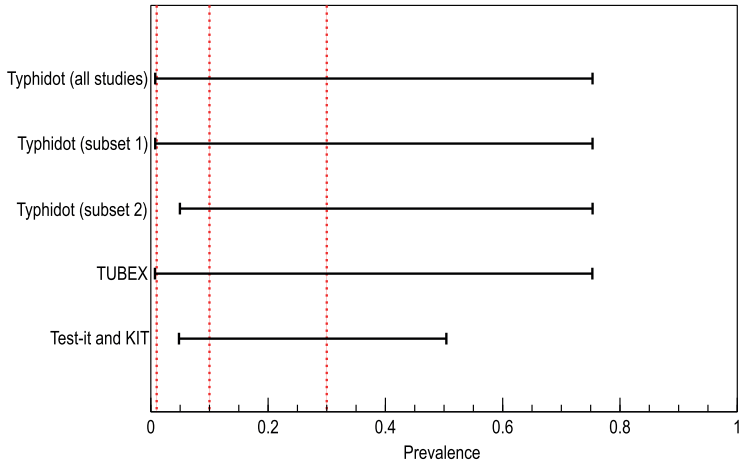


Figure A1. Prevalence ranges of meta-analyses within a target condition. *Black lines represent the prevalence range; Red dotted lines represent the pretest probabilities at 1%, 10%, and 30% selected by the review authors.*

depends on the severity of the downstream consequences of misclassifications by the diagnostic test. When these consequences are severe, it seems crucial to have a high confidence in the calculated data with pretest probabilities. In this situation the need for certainty may outweigh the need for interpretability. When there is low confidence in the calculated data by using pretest probabilities, it sometimes might seem better not to present normalized frequencies. However, not presenting normalized frequencies for the interpretation of accuracy parameters may complicate the judgement of clinicians, policy makers, and guideline boards whether to use the diagnostic test or not as well.

Rotator cuff example

The accuracy for diagnosing any rotator cuff (i.e. shoulder muscles) tears in persons for whom surgery is considered of MRI and ultrasound was assessed in a Cochrane review by Lenza et al. (2013).[2] False negative misclassifications might lead to untreated rotator cuff tears with persisting complaints (i.e. pain, limited shoulder movement), while false positive misclassifications may ultimately lead to unnecessary surgical shoulder incisions and further diagnostic testing. The downstream consequences of misclassifications are very inconvenient at least but are not affecting more serious events such as mortality. Therefore, we might accept more uncertainty from the data calculated with pretest probabilities to aid the interpretation of accuracy parameters. In Table A1, we have calculated normalized frequencies with a representative pretest probability for rotator cuff tears in symptomatic shoulders in the Japanese general population. There may be indirectness because of the population (i.e. Japanese), because of the target condition (symptomatic shoulder versus persons suspected of rotator cuff tears whom were considered for surgery), and the prevalence range of the MRI meta-analysis did not contain the pretest probability (the prevalence ranged from 50% to 96%). When the downstream consequences are considered acceptable and not severe, such indirectness might be acceptable for the interpretation of

these normalized frequencies to decide whether or not to use MRI for diagnosing rotator cuff tears in symptomatic shoulders in the Japanese population. Especially when this is the best evidence available for that particular setting.

Table A1 – *Normalized frequencies for MRI diagnosing any rotator cuff tear in symptomatic shoulders in the general Japanese population*

MRI for detecting any rotator cuff tears in a hypothetical cohort of n=1000
Summary sensitivity: 98%
Summary specificity: 79%

Pretest probability	True positive	False positive	False negative	True negative
36% (symptomatic shoulders)[1]	353	134	7	506

[1]In the Japanese general population, from Yamamoto et al. (2010)[3]

Cancerous pancreatic lesion example

In the DTA review by Best et al. (2017)[4] imaging modalities for pancreatic lesions were analyzed. Misclassifications of cancerous pancreatic lesions have severe consequences. False negative misclassifications might lead to tumor growth and early death, while false positives might lead to psychological distress, unnecessary treatment, and intervention induced complications. Different (appropriate) pretest probabilities may yield very different test performances in a somewhat wide prevalence range (see Example 3) and even small differences in false negatives might influence the decision whether or not to use a certain imaging modality in cancerous pancreatic lesions. The presented data calculated from pretest probabilities should then be interpreted at least with caution due to the severity of consequences in misclassifications. The review authors included three studies for PET diagnosing cancerous versus benign pancreatic lesions. The prevalence was 63%, 80%, and 82% in these studies and we calculated normalized frequencies with these pretest probabilities (see table A2). We also calculated normalized frequencies for the prevalence in a hospital setting (30.4%, Walter et al. [2016][5]). The prevalence range of the PET meta-analysis (63-82%) did not contain the 30.4% pretest probability. Therefore, the pooled accuracy estimates do not contain data about the test performance at the hospital setting pretest probability and there might possibly be some indirectness. It is now the question whether we believe that the normalized frequencies are a truthful (enough) representation of the test's performance considering the severe downstream consequences of misclassifications.

EXAMPLE 3 – VARYING NORMALIZED FREQUENCIES IN HIGH GRADE VESICO-URETERAL REFLUX

In the Cochrane DTA review of Shaikh et al. (2016)[6] summary sensitivity and specificity were obtained for both ultrasound and dimercaptosuccinic acid renal scans used for detecting vesicoureteral reflux. The authors assumed a pretest probability of 13% in a hypothetical cohort of n=1000 for high grade vesicoureteral reflux. We used the data of the ultrasound analysis for high grade vesicoureteral reflux to calculate the mean, median, interquartile range (IQR), and the range of the disease prevalence and calculated the accompanying

normalized frequencies. The authors were able to include 11 studies (2498 participants) in the ultrasound analysis for high grade vesicoureteral reflux. This resulted in a summary sensitivity and specificity of 59% and 79%, respectively. From the data of the analysis (test 2 on page 115 of the DTA review), the disease prevalence of the included studies for the ultrasound analysis were calculated: 8%, 9.1%, 9.3%, 15.6%, 16.6%, 17.7%, 19.5%, 20,1%, 21,2%, 23.9%, 31%

Therefore:
Range lower limit = 8%
IQR lower limit = 9.3%
Assumed = 13%
Mean = 17.5%
Median = 17.7%
IQR upper limit = 21.2%
Range upper limit = 31%

Calculating the normalized frequencies

The mean, median, interquartile range, range and the assumed pretest probability were used as pretest probabilities when calculating the normalized frequencies. The sensitivity and specificity used in these calculations will remain unchanged. With this we show that normalized frequencies vary when using different pretest probabilities determined by various methods (see Table A3).

Table A2 – *Normalized frequencies for PET diagnosing cancerous pancreatic lesions (cancerous vs. benign)*

PET for detecting pancreatic lesions (cancerous vs. benign) in a hypothetical cohort of n=1000
Summary sensitivity: 92%
Summary specificity: 65%

Pretest probability	True positive	False positive	False negative	True negative
80% (median)	736	70	64	130
30.4% (hospital setting)[1]	280	244	24	452

[1]Proportion of persons suspected of pancreatic cancer in a hospital setting diagnosed with pancreatic cancer, from Walter et al. 2016[5]

Table A3 – *Normalized frequencies for various pretest probabilities while the sensitivity and specificity remain constant*

Ultrasound for detecting high grade vesicoureteral reflux in a hypothetical cohort of n=1000
Summary sensitivity: 59%
Summary specificity: 79%

Pretest probability	True positive	False positive	False negative	True negative
8% (range lower limit)	47	193	33	727
9.3% (IQR lower limit)	55	190	38	717
13% (assumed)	77	183	53	687
17.5% (mean)	103	173	72	652
17.7% (median)	104	173	73	650
21.2% (IQR upper limit)	125	165	87	623
31% (range upper limit)	183	145	127	545

IQR: interquartile range

LITERATURE LIST

1. Wijedoru L, Mallett S, Parry CM. Rapid diagnostic tests for typhoid and paratyphoid (enteric) fever. Cochrane Database Syst Rev. 2017 May 26;5:CD008892. doi: 10.1002/14651858.CD008892.pub2. Review. PubMed PMID: 28545155; PubMed Central PMCID: PMC5458098.
2. Lenza M, Buchbinder R, Takwoingi Y, Johnston RV, Hanchard NC, Faloppa F. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. Cochrane Database Syst Rev. 2013 Sep 24;(9):CD009020. doi: 10.1002/14651858.CD009020.pub2. Review. PubMed PMID: 24065456; PubMed Central PMCID: PMC6464715.
3. Yamamoto A, Takagishi K, Osawa T, Yanagawa T, Nakajima D, Shitara H, Kobayashi T. Prevalence and risk factors of a rotator cuff tear in the general population. J Shoulder Elbow Surg. 2010 Jan;19(1):116-20. doi: 10.1016/j.jse.2009.04.006. PubMed PMID: 19540777.
4. Best LM, Rawji V, Pereira SP, Davidson BR, Gurusamy KS. Imaging modalities for characterising focal pancreatic lesions. Cochrane Database Syst Rev. 2017 Apr 17;4:CD010213. doi: 10.1002/14651858.CD010213.pub2. Review. PubMed PMID: 28415140; PubMed Central PMCID: PMC6478242.
5. Walter FM, Mills K, Mendonça SC, Abel GA, Basu B, Carroll N, Ballard S, Lancaster J, Hamilton W, Rubin GP, Emery JD. Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. Lancet Gastroenterol Hepatol. 2016 Dec;1(4):298-306. doi: 10.1016/S2468-1253(16)30079-6. Epub 2016 Oct 4. PubMed PMID: 28404200; PubMed Central PMCID: PMC6358142.
6. Shaikh N, Spingarn RB, Hum SW. Dimercaptosuccinic acid scan or ultrasound in screening for vesicoureteral reflux among children with urinary tract infections. Cochrane Database Syst Rev. 2016 Jul 5;7:CD010657. doi: 10.1002/14651858.CD010657.pub2. Review. PubMed PMID: 27378557; PubMed Central PMCID: PMC6457894

CHAPTER 9
General discussion

1. A THESIS FOR FUTURE HEALTHCARE ECOSYSTEMS

The current thesis provides an initial step into frameworks and tools to support real-world processes in a healthcare ecosystem. As we envision well-functioning ecosystems, this thesis lays a blueprint. It outlines the next steps necessary for individual organizations to pivot towards processes supporting their role and responsibilities in a broader, interconnected network of organizations – a network that will ultimately evolve into a fully realized healthcare ecosystem.

In **Chapter two** a large, overarching healthcare ecosystem was envisioned. An evidence ecosystem and learning healthcare system together have mutual beneficial associations. With well-coordinated processes, they can become symbiotic. This chapter described a model where the evidence ecosystem and the learning healthcare system are integrated into a Symbiotic Healthcare Ecosystem (SHE) framework and identified the entities within the system. Such framework could be used as a blueprint to speed up and improve complex processes in the pathway from evidence generation to implementation and learning, thereby identifying critical stakeholders, pathways, infrastructure, and products. Lots of different clinical questions arise in a SHE, many of which are not ideally suited for investigation through a randomized controlled trial. Consequently, relying solely on evaluations of care using real-world data may prove insufficient to immediately change current practice or introduce new interventions. However, real-world data can still be used to examine the claims made in the contextualized evidence and to form new hypotheses to future research. Yet, it's important to recognize that the current inefficiencies in these pathways are multifaceted, and not all delays from data generation to implementation are unnecessary. For example, we might need to accept that (re)producing primary data and rigorous developing of systematic reviews or guidelines take a certain amount of time before reaching a level of certainty deemed acceptable for integration into clinical practice. However, unnecessary delays should be prevented in the SHE. **Chapters three** to **eight** and their lessons learnt are discussed in the context of **Chapter two**, that is, in the context of a SHE.

Chapter three provided a framework to evaluate best practices. A thorough evaluation of a best practice provides the opportunity to learn and improve, as a basis of a learning healthcare organization. The framework helps to understand what important aspects of a practice in a hospital setting could be evaluated and helps to identify transferable aspects. In this chapter, it can be learned that a practice has three important categories (i.e., culture, resources, and learning) over three axes (i.e., outcomes, process, and learning). By utilizing these categories and axes within this framework, organizations gain a deeper understanding of their practices within their unique contexts, facilitating the identification of transferable elements that can benefit other healthcare practices within a SHE. Furthermore, to harness the collective learning potential within a SHE, it's imperative that information about best practices is readily shareable among participating organizations. To facilitate this knowledge exchange, a tool based on indicators outlined in the framework was provided to assist in the learning process. This enables other healthcare practices to use this as mirroring information, allowing them to reflect on and learn from it. In the context of **Chapter two**, we have learned

that practices need to be connected in a network, need to have access to infrastructure, and be able to communicate with each other. This is pivotal in fostering collaboration and driving continuous improvement across the healthcare ecosystem.

The framework presented in **Chapter four** helps to design real-world processes to let products (e.g., systematic reviews and clinical practice guidelines) continuously reflect the latest state of scientific knowledge and clinical practice. This might be achieved by a continuous maintenance cycle. Failing to maintain products over time in the SHE can lead to outdated products being used to inform clinical decisions at the point of care. The chapter addressed this responsibility in a SHE from the perspective of evidence synthesizing and guideline developing organizations. In this chapter, it was illustrated how such cyclical processes are not just updating strategies. Instead, such a process could be a broader maintenance strategy for a whole portfolio of systematic reviews or guidelines circulating within the SHE. Assessing the need for updating and performing updates is just one of the pathways in a larger maintenance strategy to continuously perform maintenance on the portfolio. For example, other management option can include withdrawal, archiving, or deferring, even when the guideline is not outdated. This framework draws parallels to clinical diagnostic test-treatment strategies, where multiple treatment options are available, and the choice of treatment is guided on information obtained from diagnostic tests. Subsequent monitoring at specific intervals thereafter triggers re-assessments. Alternative test and monitoring options can be tailored to suit various organizational requirements and contexts. Designing a maintenance strategy that is capable of adequately maintaining a portfolio could increases trust in the SHE. Trust is very important for the end-users who need to rely on accurate, up-to-date information for clinical decision-making, necessitating products that consistently reflect the latest state of scientific knowledge and clinical practice. Obviously, maintaining products within a SHE is only meaningful when these products successfully reach the end-user. Therefore, the organizations developing products and the end-users need to be connected in a network (such as described in **Chapter two**) with an infrastructure to ensure fast communication (e.g., to exchange information and products) with each other.

Chapter five continued to extend the framework presented in **Chapter four** with adding priority-setting assessments to maintenance strategies. Priority-setting is essential when dealing with a system in which resources are limited. This chapter illustrated how the outcome level and position of the priority-setting assessments in a maintenance strategy influences its performance. Priority-setting assessment outcomes will steer in which systematic reviews or guidelines resources need to be invested in. **Chapter four** and **five** together demonstrate that priority-setting is a different construct than assessing the need for updating. The need for updating concerns testing for indications (i.e., staging) whether an update is needed. Priority-setting, however, tries to identify a set or individual items which have more priority over another set or individual items. However, if the evidence synthesizing and evidence translating organizations do not prioritize their products for maintenance, they fail their responsibility in a SHE. This might lead to outdated products being used to make healthcare decisions at the point of care, affecting both patients and clinicians. Placing **Chapters four** and **five** in the context of **Chapter two**, we have learned that organizations may need to adjust their current internal processes and pivot towards such cyclical maintenance processes that respect resource limitations and meet the responsibility of continuously letting their products

reflect the latest state of scientific knowledge and clinical practice. With **Chapters four** and **five** together, systematic review or guideline developing organizations now have a framework to do so. However, they may require additional guidance to facilitate and conduct their real-world priority-setting assessments from this perspective.

Chapter six builds upon previous chapters. From **Chapters two**, **four**, and **five**, it was learned that these priority-setting assessments and their supporting tools should be flexible enough to meet the varying unique contexts and needs the organizations operate in. **Chapter six** provided guidance and a tool to operationalize real-world priority-setting assessments. The presented frame of reference aimed to help to understand and structure new and existing priority-setting assessments. The accompanied RE-weighted Priority-Setting (REPS)-tool is a flexible tool to determine the level of priority in such assessments. This chapter illustrated how priority-setting assessments consist of three main components chained together: process components (procedural steps), function components (a mechanism or set of rules to define the level of priority), and an outcome component (final outcome of the priority-setting assessment). From this, it can be understood how the REPS-tool is a function component in a priority-setting assessment, using the re-weighted range voting mechanism to define the level of priority. Any procedural component before using the REPS-tool can be used as long as the procedures lead to the correct input format for the tool (i.e., one score per participant per item). This is important, as organizations in a SHE operate in different contexts, with different stakeholders, have variable resources, and have different needs which may need different processes. Any item can be prioritized in the REPS-tool, although examples in this chapter concern guideline key questions only. The output of the REPS-tool can immediately be used as the outcome of the priority-setting assessment, or any process component after using the tool can lead to the definitive outcome of the assessment. The re-weighted range voting mechanism furthermore ensured a proportional representation, while its adjusted mechanism allows for a disproportional representation. This is important when imbalances in the representation of the priority-setting assessment's participants underexpose important perspectives from less represented participants. The REPS-tool's output is reproducible to ensure transparency and trust in its output. The tool can be utilized in a priority-setting assessment in a maintenance strategy, such as described in **Chapter five**, aiding organizations in a SHE to determine which of their products have the most priority to spend their limited resources on. This could further support real-world cyclical maintenance processes while acknowledging resource constraints.

Once new topics are prioritized for development or existing topics for updating, it is more resource efficient to (re)use each other's efforts. This requires coordination and communication between entities and their actors in a SHE. **Chapter seven** described a real-world example of how the evidence synthesizing and translating entities were bridged. In this connection, spent efforts in systematic review development were reused in the development of a clinical practice guideline. The chapter demonstrates that transparent and complete reporting in one entity is important to reuse the efforts of producing a quality product in another entity. Reusing each other's efforts in a SHE is important to prevent unnecessary duplicate work and thus be more efficient with available resources. Observing **Chapter seven** through the lens of **Chapter two**, such connections require trust and communication between evidence synthesizers and guideline developers in a sense that the systematic review is developed rigorously while its information is packaged in a form that is relevant, easily sharable,

accessible, and usable by the evidence translator (e.g., the guideline developer).

However, communication within a SHE that only results in sharable information and products reaching the end-user is not enough. One of the most important steps in a SHE is communicating data and products in an interpretable manner for their end-users. **Chapter eight** describes how the (summary) diagnostic accuracy parameters can be translated through a hypothetical cohort to interpret these parameters. This, however, requires the selection of a pretest probability, which influences the number of (mis)classifications in the hypothetical cohort. The chapter gathers methods how Cochrane Diagnostic Test Accuracy Reviews have selected their pretest probabilities and discusses some considerations when selecting a supposedly appropriate pretest probability to communicate accuracy findings to the end-users (e.g., clinicians). This chapter argues that some indirectness in the translation of diagnostic accuracy parameters is unavoidable this way. The indirectness concerns the (summary) sensitivity and specificity being fixed over a range of representative pretest probabilities in the hypothetical cohort, where in reality they may vary from the (summary) sensitivity and specificity. In the context of a SHE, it may nonetheless be important to try and communicate findings in an interpretable manner to end-users despite some indirectness rather than communicating difficult-to-interpret parameters. Infrastructure needs to be in place so that end-users and producers of products are connected in a SHE, making sure that data and products are sharable with end-users. Co-creation in the SHE could support the relevant packaging of information to increase usability and interpretability for the end-user. By doing so, products in the SHE could match with the need of end-users to make informed (clinical) decisions based on (translated) products.

Re-iterating the 2013 Institute of Medicine's quote, we need to make sure a future SHE has a clear plan with contributing goals:

> *"If home building were like healthcare, carpenters, electricians, and plumbers each would work with different blueprints with very little coordination."* [1]

In this analogy, our carpenters, electricians, and plumbers (e.g., clinicians, patients, researchers, synthesizers, translators, implementors) need to have the same blueprint (i.e., a well-defined strategy with contributing goals) and need to determine who does what and when (i.e., coordinated). However, through this thesis, we have learned that actors involved in "home building" also needs to be connected, communicating, and trusted. We extend this analogy by adding that these actors also need to be able to find and know each other to be connected, need to talk to each other in same language to be communicating, and need to perform their role and responsibilities according to construction standards on a regular basis to be trusted.

2. THE SYMBIOTIC HEALTHCARE ECOSYSTEM

2.1 A WELL-DEFINED STRATEGY WITH CONTRIBUTING GOALS IN THE SYSTEM

Visualizing a new or existing network of organizations by using the SHE-framework (**Chapter two**) helps in creating a blueprint for all the participating actors in the healthcare network. This increases the understanding that the learning healthcare system[2-6] and the evidence ecosystem[6-10] are not separated systems. Instead, they have mutual beneficial associations, making the systems symbiotically interact with each other as if they are one single, larger system. The overarching goal of the SHE system is to consistently deliver the best care to patients. This overarching goal implies that the system should continuously evaluate, learn, and improve upon itself over time to continuously keep providing the best care. This commitment to learning and improving goes for both the learning healthcare system-part and the evidence ecosystem-part of a SHE.

Although evaluation (analyzing) and learning are already prominent elements in the learning healthcare system,[2, 3] the evidence ecosystem does need to learn and improve as well. Over time we already have seen developments in, for example, experimental research design,[11] in introducing reporting guidelines (e.g., CONSORT,[12] PRISMA[13]), in systematic review methodology (the online Cochrane Handbook's is currently on its 6th version [v6.4][14]), in methods correcting for confounding in observational studies,[15] and in network meta-analyses,[16] indicating that the evidence ecosystem-part also seems well capable of learning and improving to support delivering the best care to patients. The SHEs overarching goal, however, may never be reached despite all improvements in both sub-systems. In reality, it might be a never-ending cycle of evaluation, learning, and improvement.

Interpreting the overarching goal as an ideal, the system, as in every ecosystem, may try to obtain some form of homeostasis instead. That is, a balance in preferred positive outcomes and (un)acceptable negative outcomes being considered important at that point in time as the best possible care. This balance of positive and negative outcomes may be reflected in a best practice for healthcare provision, where all efforts by entities in a SHE are combined to provide the best available care at that moment in time. Once a practice is identified as a best practice, we can use the framework and tool described in **Chapter three** to evaluate this practice on the culture, resources, and learning domains. This information can thereafter be used to learn from by other practices once shared across the SHE. Thus, positive and negative outcomes of the delivered best available care may be measured at the point of care in the learning healthcare system-part of a SHE by measuring patient and disease relevant outcomes. However, outcomes also need to be interpreted in a broader sense (e.g., resource costs, healthcare provider well-being, culture, ability to learn). Changing perspectives on preferred positive outcomes and (un)acceptable negative outcomes, or their balance, might disrupt the system and cause adaptations (e.g., significantly altering processes and procedures within and between organizations in a SHE) to thereafter reach a form of homeostasis again. These disruptions can depend on (societal) values and ethics over time. As a society, we currently may not find expensive medication relative to small gains

acceptable for example, although, this might be different for other societies and cultures or even change over time. This accentuates how global evidence in the evidence ecosystem is translated to the local context for application at the local point of care (e.g., in a specific society with a specific culture and healthcare system) in a SHE.

The difference in levels of the two sub-systems in a SHE underscores a helpful benefit. Since global evidence is translated to a local context and healthcare is provided in a local context, all data gathered in this local context could be used to evaluate whether the claims made in the global evidence substantiating the translated products actually may hold true in the local practice. **Chapter two** briefly discussed how care could be evaluated by simulating target trials with observational data[17, 18] gathered from routine care to provide real-world evidence. Even though there still may be some difficulties in this methodology[18] (e.g., residual confounding), it might currently be the best method for causal inference using observational data.[17] With that, it might provide an accurate enough estimate to understand whether claims may hold true in the localized context.

Goals of individual organizations should contribute to the overarching goal of the ecosystem, even when they not directly interact with care provision in a 'patient-clinician interaction' sense. Individual goals may concern setting (pre)conditions that influence the system's ability to reach its overarching goal. However, organizational objectives do not necessarily need to align, as they may coincide within an ecosystem due to differences in roles and responsibilities. For example, guideline developers might aim to describe the best care in their guidelines regardless of its restitution or costs, while governmental policymakers and healthcare insurers within the healthcare ecosystem seek to manage healthcare expenses. Even though the overarching goal may be shared (i.e., provide the best care), the organizations have different objectives interacting with the system's ability to reach its overarching goal, whereafter homeostasis may be found: providing the best available care while taking into account both restitution and available resources. Therefore, the system would find homeostasis through contributing goals rather than striving for shared or aligned goals, despite the likelihood that the overarching goal is probably shared among all participating organizations and entities in the healthcare ecosystem.

2.2 COMMUNICATION AND COORDINATION IN A CONNECTED SYSTEM

A first step for a connected and coordinated network is identifying and being explicit about each participating organization's role and responsibilities in the SHE. From there, it needs to be derived which entities are capable to perform certain actions and develop certain products. Thereafter, it could be asked '*for who?*' and '*who to involve?*' to identify important links between organizations fostering co-operation, co-creation, and the identification of (end-)users of the organization's products. The SHE-framework in **Chapter two** helps in identifying important links and relations. Different organizations and stakeholders in the evidence ecosystem-part of a SHE develop various products. For example, the research entity produces primary data and research reports, the evidence synthesis community produces (systematic) reviews, and the evidence translating community may produce clinical practice guidelines. These efforts, however, are currently performed siloed.[19]

In **Chapter seven** we have provided a real-world example of how a bridge between evidence synthesizers and guideline developers had benefits. The developed guideline segment reused the efforts in two Cochrane reviews[20, 21] produced in the evidence synthesis entity. Reusing these efforts enabled the identification of three primary studies[22-24] for a subgroup of patients with a rare disease, instead of re-doing the same search strategy and literature selection. It was fortunate enough that, in the absence of a functioning evidence ecosystem, these high-quality Cochrane reviews were found and that some of their data was accessible and helpful. For effective coordination within a SHE, governing bodies would probably need to establish minimal acceptable practices for the development systematic reviews, such as adhering to the PRISMA-statement,[13] and ensuring data availability and accessibility. Similarly, guidelines can be guided by standards like AGREE-II.[25]. Additionally, organizations in the evidence synthesizing and translating entities would need to establish consensus for reporting their products in formats that are exchangeable and usable. Besides reusing each other's data and products, both entities could align their efforts in a SHE. Here, the guideline developing organization involves the evidence synthesizing entity in their planning of developing systematic review-substantiated guideline segments (*'who to involve?'*), and the evidence synthesizing entity produces commissioned systematic reviews for the evidence translating entity (*'for who?'*). This aligns efforts and maximizes the potential impact within the SHE.

Adhering to reporting guidelines in a SHE is a first step. It is also important to consider how data in various products can be packaged and exchanged. Products, such as clinical practice guidelines and systematic reviews, are being developed from smaller pieces of information. When this information is compartmentalized, information blocks emerge in a SHE. For example, a guideline can be compartmentalized into multiple guideline segments. These information blocks can be defined and developed in such a way that they are immediately relevant for other organizations in the system. A marketplace-like database on a digital platform could enable the exchange of such blocks across a SHE (see **Chapter two**). Figure 1 shows how a collection of information blocks resembles some of the products in a SHE and how they are associated with each other. As shown in Figure 1, a clinical practice guideline segment can be compartmentalized in blocks containing the key question, a systematic review, deliberations or structured evidence-to-decision,[26] a body of recommendations, a reference list, and identified knowledge gaps that prevent (stronger) clinical policies.

Some information blocks may encompass multiple sub-blocks, containing smaller pieces of information. For example, the body of recommendations information block. This block encompasses all recommendations and may include grouped blocks detailing the recommendation, its rationale, as well as its strength and/ or direction (see Figure 2). Compartmentalizing such data and storing it in a database could enable an effortless overview of the information in clinical practice guidelines, including recommendations, their rationale, and their strength and direction in the organization's portfolio. For the organization it is helpful for structuring de data stored in their database. Such structure provides several other benefits. It also offers a structure for database research. For example, the organization can then easily query the database for the number of strong or weak recommendations in their portfolio. It may also offer some freedom in communicating the data to their end-user in clinical practice. When compartmentalized, there is freedom to present the information

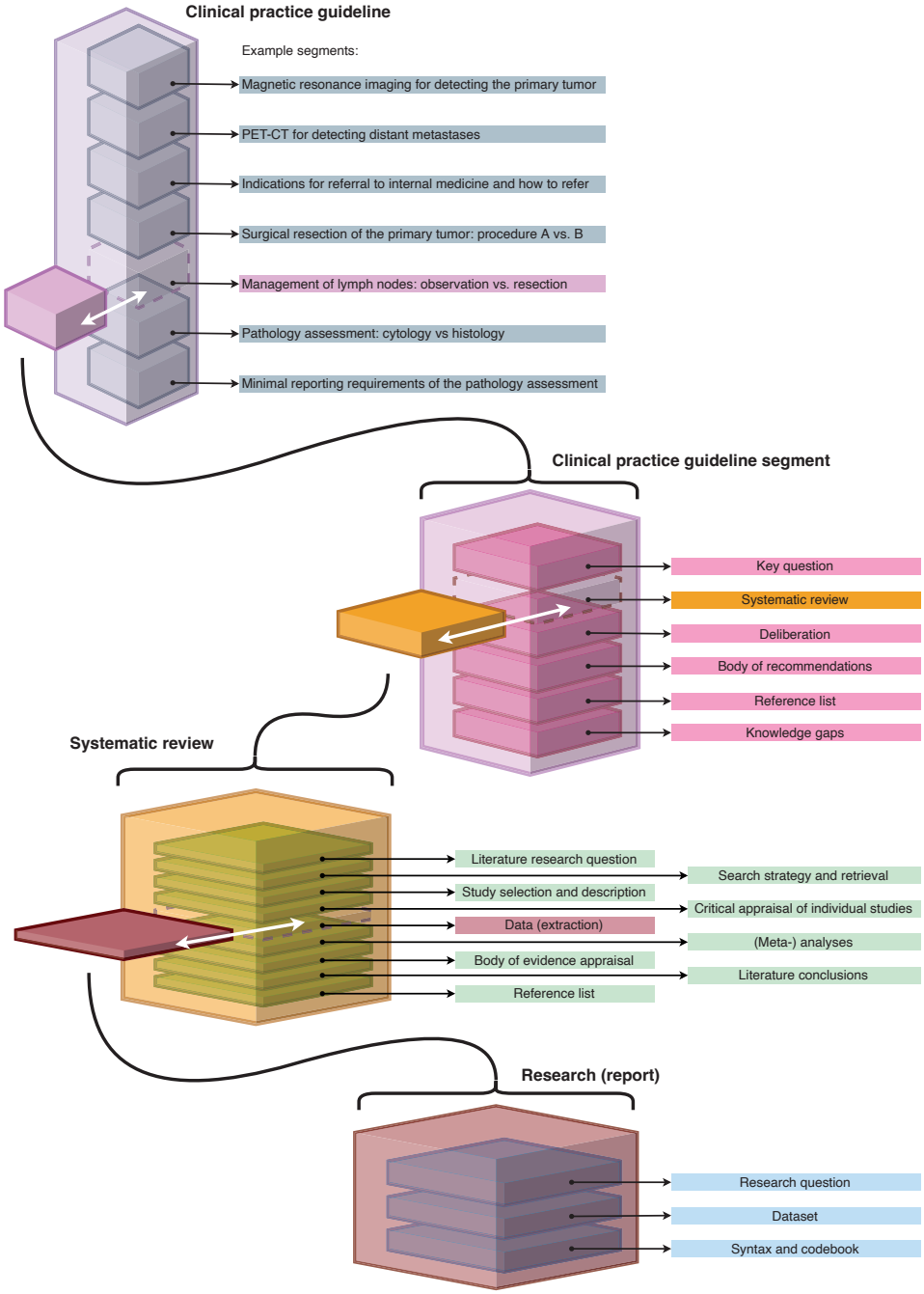


Figure 1. A visual representation of some possible compartmentalized information blocks in a symbiotic healthcare ecosystem. *Both the product and individual information blocks might be exchanged in a symbiotic healthcare ecosystem.*

blocks in predefined user-friendly formats representing the clinical practice guideline. When user-testing over time indicates that there are different or evolving preferences, a new format could be developed to add, remove or re-shuffle the displayed information blocks. This can even be user-directed in the future, where users determine for themselves how their guidelines are displayed in a format relevant and usable to them, whether it is displayed on a webpage or in electronic health record software. Overall, structuring information blocks facilitates seamless database storage, organization, and adaptable presentation formats to meet evolving user preferences.

Compartmentalization furthermore offers opportunities to exchange information blocks across a SHE. For a clinical practice guideline, for example, several information blocks within a clinical practice guideline segment can be shared or received (see Figure 3). Sharing key questions with other guideline developing organizations may communicate which guideline segments are already (being) developed. This could increase coordination and might prevent unnecessary duplicate efforts across the SHE if the guideline segment (or evidence synthesis) is thereafter also findable, accessible, interoperable, reusable, and trustable[27] for other organizations. Key questions should also be shared with the evidence synthesizing community for the development of systematic reviews substantiating clinical practice guideline segments. Systematic reviews should be exchanged and effortlessly inserted in clinical practice guideline segments when packaged relevantly. It is arguable whether the deliberations and body of recommendations information blocks in a guideline segment are

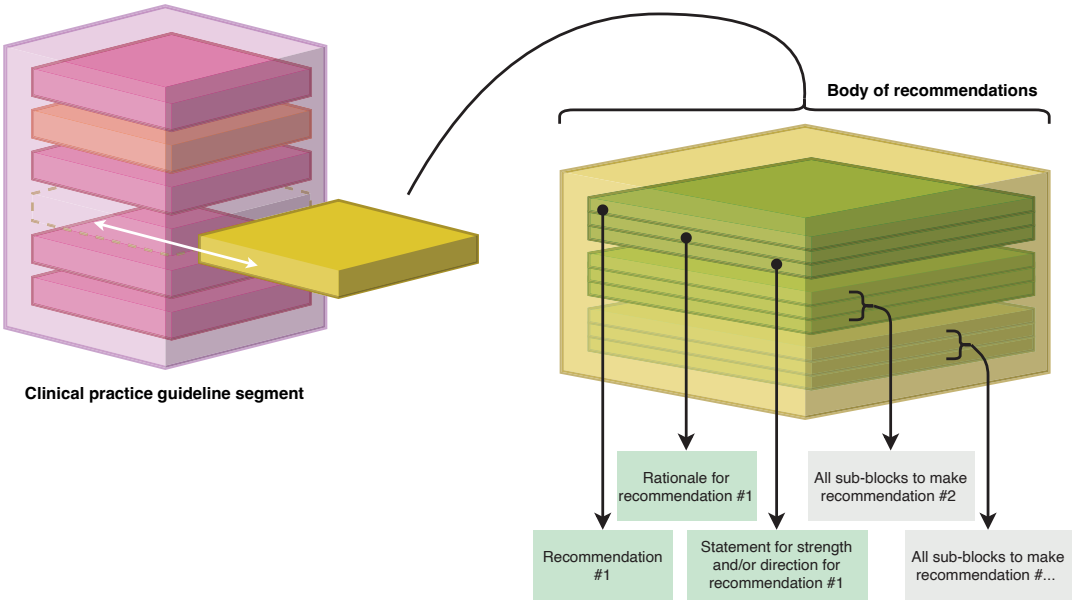


Figure 2. Sub-blocks in the body of recommendations information block. *The sub-blocks contain information regarding the recommendation, its rationale, and the strength and/ or direction of the recommendation. The body of recommendations contain all recommendations in the clinical practice guideline segment, from the first (#1) up to any other number (#...) of recommendations developed in this guideline segment.*

directly usable for exchange in an international context, as they may be tailored to the local healthcare system or situations specific for a single country. However, there are frameworks available to adapt existing clinical practice guidelines.[28, 29] The deliberation and body of recommendation blocks should be shared with the end-user in the clinical field. During the development of the guideline segment, knowledge gaps can be identified that prevent making (stronger) clinical healthcare policies. Sharing these evidence gaps with the primary research community aligns primary research with existing and the most urgent knowledge gaps for clinical practice. Another product in a SHE that can be compartmentalized is the systematic review, shown in Figure 1. It is an information block in a clinical practice guideline segment but can be produced in the evidence synthesis entity for exchange with the guideline developing community in the evidence translating entity (see Figures 3 and 4). When the information and data in systematic review products are compartmentalized completely, transparently, and are available and accessible, their information also becomes reusable for other communities in the healthcare ecosystem. **Chapter seven** showed real-world benefits of transparent and complete reporting of systematic reviews for clinical practice guideline development, where the '*search strategy and retrieval*', '*study selection and description*', '*data (extraction)*', and *'reference list'* information blocks of the systematic reviews were reused.

Effectively communicating compartmentalized data and information in a coordinated manner across the system requires more than just adherence to reporting guidelines. There is a fundamental need standardization and harmonization of data. To enable the exchange of information, digital information blocks within products need to be findable, accessible, interoperable, reusable, and trustable[27] for all stakeholders. One potential solution particularly relevant for a SHE is the Fast Healthcare Interoperability Resources (FHIR), designed to facilitate data exchange across various healthcare, healthcare-related, public health, and research settings.[30] FHIR could make data and databases interoperable. However, healthcare and healthcare-related data only may concern the point of care with its associated actors, organizations, and databases (e.g., clinical data in electronic patient records, registries) from the learning healthcare sub-system of a SHE. Such a common language probably needs extensions to cover information produced and exchanged elsewhere in a SHE as well. Projects like EMBonFHIR aim to address this gap by extending FHIR to record and exchange data generated in primary studies and evidence synthesis,[31] thereby enhancing interoperability within the evidence ecosystem sub-system in a SHE. Similar to the concept of compartmentalizing information in primary research and evidence syntheses presented in Figures 1 to 3, EBMonFHIR describes distinct and identifiable elements in EBM-products to support communicating scientific evidence.[31, 32] FHIR seems to be "close to the metal", meaning it seems to capture data on a "low-level" in fundamental information blocks. This enables exchange of data that evidence synthesizers need from primary studies, and data that guideline developers need from systematic reviews. Such "close to the metal" initiatives for primary studies and evidence syntheses might be extremely useful in a SHE, as the compartmentalized information represents relevant and reusable information (sub)blocks and even smaller data elements. Storing study data in this structured manner could potentially offer a more efficient streamlining of the process compared to what was discussed in **Chapter seven**. Then, all the study data would be compartmentalized and findable using a simple database query returning data about the requested elements (e.g., patient / intervention / comparison / outcome elements). This could ultimately reduce the need for manual data-

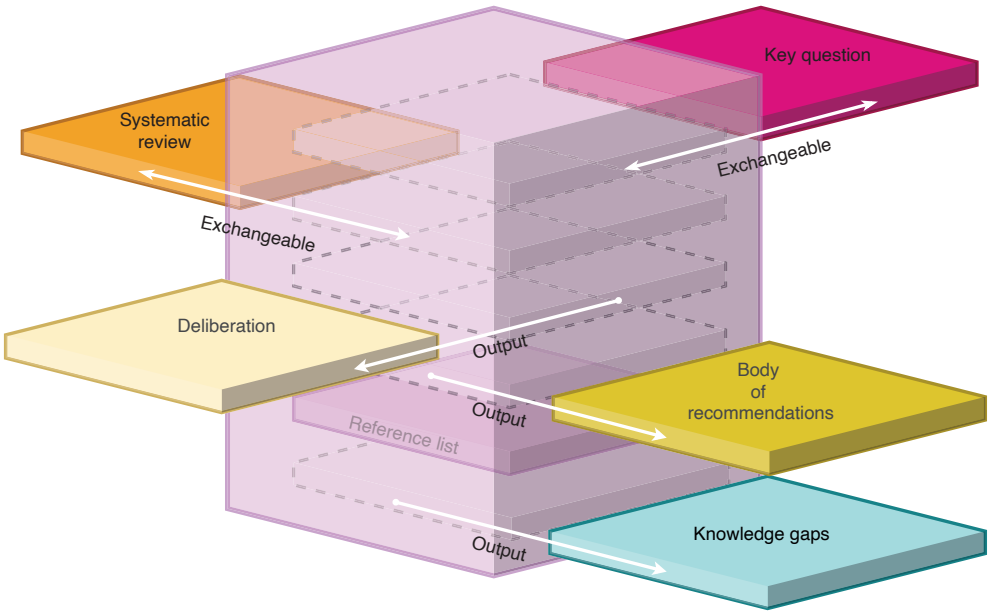


Figure 3. Potentially sharable information blocks in a clinical practice guideline segment. *Key questions can be distributed to prevent duplicate efforts. Systematic reviews could be received from the evidence synthesis community or shared with other guideline developing organizations. Deliberations and the body of recommendations could be shared with the clinical field. Knowledge gaps could be shared with the primary research entity.*

extraction from primary study reports in the future. Furthermore, the CPG-on-EBMonFHIR[33] project could provide a "close to the metal" option for the evidence translation entity in a SHE. This makes the recommendations and deliberations from guideline segments immediately sharable with the healthcare providers, which could be linked with patient characteristics to both show patient-relevant recommendations and enable guideline adherence monitoring in clinical practice.[34]

Infrastructure and collaboration seem essential for the flow of evidence to clinical care.[6] The benefits described in **Chapter seven** can be utilized constantly in a future SHE if compartmentalized data can be efficiently exchanged between entities using a digital infrastructure. Web-apps in the digital infrastructure, like RevMan Web,[35] MAGICapp,[36] and GRADEpro GDT,[37] could enforce formats for producing and reporting evidence syntheses and guidelines, all while capturing and storing data using a common language such as FHIR. In fact, the MAGICapp seems to be planning to implement EBMonFHIR and CPGonFHIR standards and is part of these projects' working groups.[33] Therefore, such web-apps might be in a unique position to become a marketplace-like platform for the coordinated exchange of compartmentalized data and products in a SHE. For example, evidence gaps from guidelines can be shared with the primary research entity. The resulting new primary data can then be synthesized using RevMan Web and embedded in guidelines with the MAGICapp or GRADEpro GDT. The guideline can then directly be used in the local clinical

practice, integrated in electronic health record software and clinical decision aids using the same common language.

Although using an infrastructure for exchanging information between entities and reusing each other's efforts is important, ensuring that products reach end-users in the SHE is equally vital. In some cases the end-users could be patients (e.g., for patient information, decision aids) while in other cases this could be clinicians (e.g., for clinical practice guidelines) or even governmental policymakers (e.g., for policy briefs). When disseminating such products to the end-user, it is important that the information is interpretable for their application. In fact, if the data is uninterpretable for its end-user, all efforts in developing and exchanging products in a SHE might be in vain as these products cannot be understood or applied. Recognizing the importance of the interpretability of the products in the system (e.g., see **Chapter eight** for communicating diagnostic test accuracy parameters), different end-users may require tailored outputs, although the underlying data may be the same. For example, synthesized data may be communicated in the form of a systematic review report for researchers and guideline developers, while it could also be presented as a policy brief for policy makers, or as a clinical decision support product for patients. Co-creation could aid the production of such tailored outputs for end-users[19] and seems important in the SHE (recall '*who to involve?*'). When patients and clinicians co-develop clinical practice guidelines (e.g., by defining relevant problems and questions), for example, these guidelines will be more relevant for the point of care. Another important aspect of this co-creation of guidelines lies in identifying knowledge gaps and selecting relevant outcomes. When patients and clinicians determine relevant topics and questions for a guideline to provide healthcare policies, any identified knowledge gap probably immediately becomes relevant for the point of care. The knowledge gaps prevent evidence translators from making (stronger) healthcare policies based on the available evidence. Thus, the knowledge gap is immediately relevant for the primary research entity, as they guide further research to fill the identified gaps and contribute to evidence-based clinical practice through guidelines (see Figure 4). In such a cycle, new primary evidence may traditionally take a long time to find its way to clinical practice through guidelines. It is important to keep in mind that not every time lag is undesirable to ensure safety and efficacy[38] in a SHE. It may take some time to form an evidence base to make strong(er) recommendations for clinical healthcare policies. Studies need to be large enough, well-conducted, and probably even replicated to gain enough certainty in an evidence synthesis for a strong clinical recommendation. The challenge is to understand which time lags are unnecessary[38] and how a SHE could contribute to an as-soon-as-responsible transfer of new developments into clinical care. Therefore, it is crucial to acknowledge and accept necessary time lags while striving for timely implementation of new developments in clinical practice within the SHE.

2.3 TRUST IN THE SYSTEM

Trust within a SHE is two-fold for users of products. First, users need to be certain that the products they use are trustworthy. This prevents spending time and effort on performing quality appraisals for each and every product they consult (e.g., guidelines and systematic reviews). Second, users need to be certain that the products they use are maintained to reflect the latest state of scientific knowledge and clinical practice over time as well as possible.

In **Chapter two** we discussed that there are many methods and practices available to develop trustworthy products. For guideline development there are essential domains,[25] complete handbooks by the Scottish Intercollegiate Guidelines Network and the National Institute for Health and Care Excellence,[39, 40] and even a framework to adapt existing guidelines.[29] The GRADE working group provides access to a handbook to assess the certainty in a body of evidence in systematic reviews and guidelines,[41] together with ongoing publications in scientific peer-reviewed journals. Furthermore, a framework to transparently report considerations leading to guideline recommendations is provided by the GRADE working group.[26] For the evidence synthesis entity, extensive handbooks for developing intervention[42] and diagnostic test accuracy[43] systematic reviews are available, besides guidance for rapid reviews.[44] Extensive guidance is provided by the COSMIN Initiative for systematic reviews of patient reported outcome measures.[45-47] Besides sufficient information on how to produce systematic reviews, reporting guidelines for systematic reviews are furthermore available with the PRISMA statement[13] and all of its extensions. Such reporting guidelines also exist for the primary research entity with the CONSORT statement[12] for randomized trials, although there are many more reporting guidelines found through the EQUATOR network.[48] Furthermore, core outcome sets are available for many clinical fields at the COMET initiative.[49] As described, there are numerous resources and guidelines readily accessible for developing trustworthy products within the system. To trust each other's products, it might be important to determine which practices are minimally acceptable to develop products in the SHE. Here, it is wise to acknowledge variations in available resources across the system. As rigorous processes result in transparent gold-standard products, the development of such products may also cost a lot of resources. Such resources are scarce for organizations across the system and setting minimally acceptable practices can ensure that products remain trustworthy and transparent for their use.

Besides ensuring trust by using rigorous methods resulting in trustworthy products, the second factor is trusting these products over time as new developments and new evidence emerge. Organizations have the responsibility of ensuring that their products reflect the latest state of scientific knowledge and clinical practice as well as possible (**Chapters four** to **six**). Where published strategies mainly seem to focus on whether or not to update,[50-55] the framework in **chapter four** showed that multiple options are available in a maintenance strategy besides (not) updating, even when the product is not outdated. Some products may need updating, others are still useful but get archived for reference, and yet other products need to get withdrawn from the SHE. Using the framework, designers might design or tailor a maintenance strategy specific to their needs and context in a SHE, as it does not impose specific tests, criteria, or indications to the designer. This is important in a SHE, as different organizations may operate in different ([inter]national) contexts, have different needs, and have different resources available. For example, archiving guideline segments might be important in the context of one organization, but not in the context of another. Although the idea that partial updating guidelines makes more sense is a decade old,[50] compartmentalizing products in combination with the POMBYTT framework now enables the maintenance of an entire collection of guideline segments representing multiple clinical practice guidelines as displayed in Figure 4.

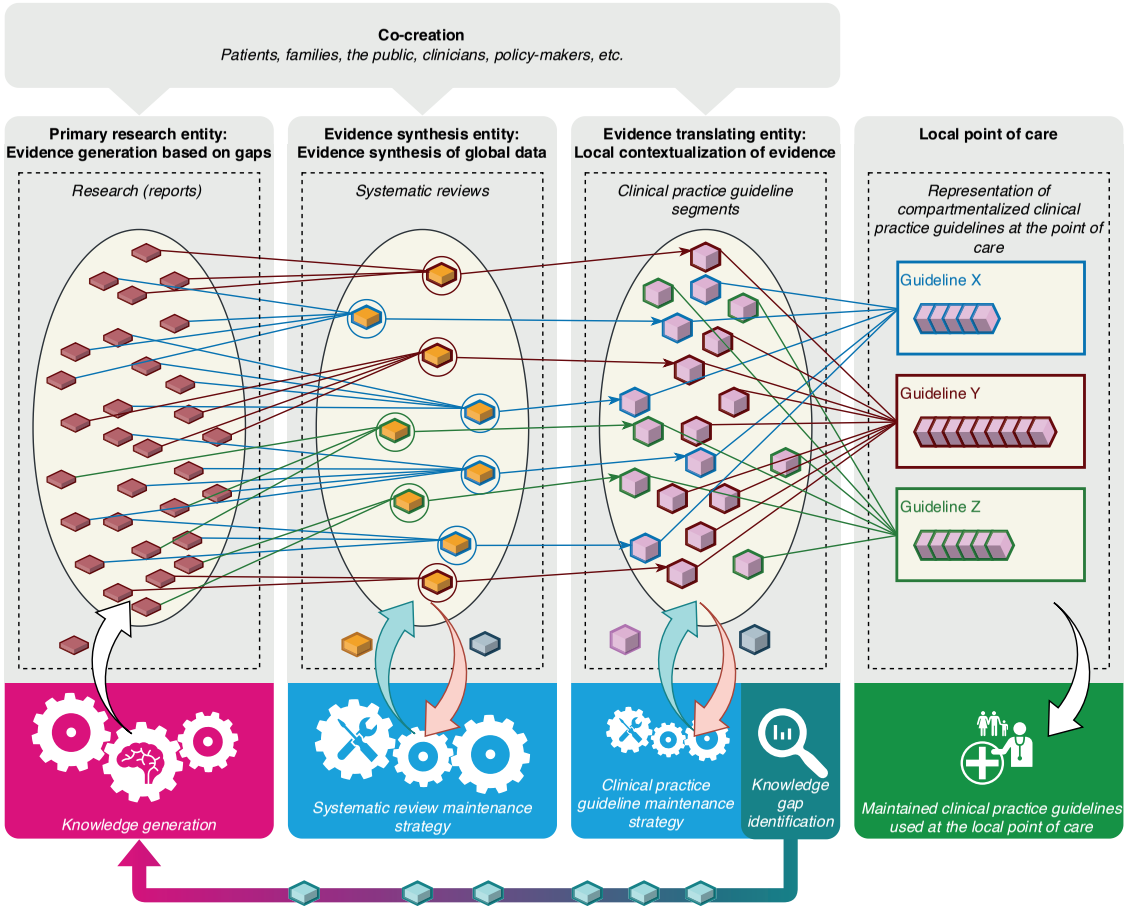


Figure 4. Sharing compartmentalized, relevant information between entities in a symbiotic healthcare ecosystem. *Knowledge is generated in a format relevant for the evidence synthesis community (red blocks). Evidence is synthesized in a format relevant for the evidence translating community (orange blocks). The evidence translating community identifies knowledge gaps (turquoise blocks) from their clinical practice guidelines (pink blocks) and communicates these with the primary research community. Both the evidence synthesis and evidence translating entity have their maintenance strategies in place to maintain their products circulating within the symbiotic healthcare ecosystem. Products are assessed in a maintenance strategy (grey blocks) and return (green and red arrows) as maintained as possible (the returning orange and pink blocks) in the portfolio. This ensures that there are maintained healthcare policies at the point of care.*

The framework in **Chapter four** furthermore allows for cyclical maintenance strategies when desirable. Instead of a fixed time interval triggering a management option (e.g., updating after three years), the framework guides designers to selecting an interval that triggers re-assessment of outdatedness and indications for management options. This way, the framework might be used to design a continuous strategy similar to 'living systematic reviews'[56] or 'living clinical practice guidelines'.[57, 58] In the latter, recommendations in a single guideline are assigned a living status.[56, 57] Literature is continuously monitored, and recommendations are updated as soon as new evidence becomes available.[56, 57] However, continuously monitoring may be very resource intensive and adding new literature might not immediately translate to a change of the body of recommendations (e.g., strength, direction, addition, or removal). The POMBYTT framework provides designers with the opportunity to explicitly think about such elements and consequences for their maintenance strategy. Thus, instead of immediately triggering an update as new evidence becomes available, a strategy designed with the POMBYTT framework tests whether an update (or any other management option) is indicated. This could prevent spending resources on updates that (likely) do not change recommendations and thus clinical practice. Organizations may select tests, indications, and management options that suit their needs and their available resources for a maintenance strategy. This way, organizations can meet their responsibility of maintaining their circulating products in a SHE at their own capacity.

When there are no resource limitations, the entire portfolio can always and continuously be maintained without constraints. However, organizations probably always have resource limitations which requires prioritization in some form. **Chapter five** extended the POMBYTT framework described in **chapter four** by providing concepts of priority-setting in maintenance strategies. Priority-setting is important, as an increasing number of products within a maintenance strategy may cause increasing resource utilization. Furthermore, there are limited resources for performing certain management options in the maintenance strategy (e.g., updating). Priority-setting is then needed to determine which products have enough priority to enter the maintenance strategy and/or have enough priority for the selected management option given the available resources. If evidence synthesizing and evidence translating organizations fail to prioritize their products for maintenance, they might fail their responsibility in a SHE. This leads to outdated products being used in a SHE to make healthcare decisions at the point of care. **Chapter five** therefore provides guidance to help organizations understand priority-setting assessments and spend their available resources for maintenance as well as possible. This intends to support the responsibility of the guideline or systematic review developing organization is a SHE, where they try to keep their products as maintained as possible and thus sparking trust in the system.

Together, **chapter four** and **chapter five** communicate an important difference between assessing the need for a specific management option (e.g., assessing the 'need for updating'[51, 53-55]) and assessing priority. This difference is important in a SHE, as not all products in need for updating may actually receive an update due to resource limitations. This distinction also seems important for clinicians, patients, and patient representatives as they may be involved during developing, maintenance, and priority-setting processes. To aid guideline developing organizations with performing priority-setting assessments in their maintenance strategies, the RE-Weighted Priority-Setting (REPS-)tool was developed and presented in **chapter six**.

The chapter first provided a frame of reference for priority-setting assessments, identifying three main components: the process component, the function component, and the outcome component. Using this frame of reference, priority-setting assessments can be structured and the role of the REPS-tool becomes clear. The REPS-tool was specifically designed to have a ranked continuous output to be used with a limit of capacity (see **Chapter five**). This means that the tool's output can be aligned to the available resources for management (e.g., updating) after priority-setting, if desirable. This can be important in a SHE, as some organizations may have more resources than others for management and the output can now be aligned with the organization's capacity. **Chapter six** furthermore introduces the (decay-adjusted) re-weighted range voting mechanism[59] in clinical practice guideline prioritization, where other priority-setting tools seem to sum or average.[60-62] The re-weighted range voting mechanism has a proportional representation characteristic,[59, 63] however with an adjustment, the mechanism achieves a more disproportional representation.[59] The latter is essential for multidisciplinary priority-setting in a SHE where important underrepresented views may be boosted in the priority-setting assessment. Each step in the REPS-tool is reproducible and leads to the same output when using the same input and tool parameters every time. The REPS-tool furthermore does not impose any procedural steps to designers of priority-setting assessments. The designer is free to select priority indicators, adjust the score range, prioritize other items (e.g., outcomes, recommendations, knowledge gaps), or add any other procedural step as long as the required input format for the tool is met (i.e., one score per item per person). This enables tailoring of the priority-assessment to the required context and is important because, for example, priority indicators may not be predictive in different contexts.[64, 65] As the context is important, organizations may also choose not to use the REPS-tool and, instead, develop their own tool or use another existing tool. One of such tools could be the UpPriority tool[61, 62] for clinical practice guidelines or the SPARK tool for systematic reviews.[60] The frame of reference, described in **Chapter six**, aids designers of priority-setting assessments in appreciating (published) tools and how they would fit in their own priority-setting assessment. Together, **Chapters four** to **six** enable organizations to maintain their circulating products within a SHE to reflect the latest state of scientific knowledge and clinical practice over time as well as possible. Such continuous, transparent, and reproducible processes can spark trust in how and which products within a SHE are maintained.

3. CONCLUDING REMARKS

This thesis presented the operational aspects of a future healthcare ecosystem, emphasizing the importance of communication within and between entities, maintenance, and evidence translation. The sustainability of the proposed healthcare ecosystem is dependent on compartmentalizing data for exchange and reusability, the underlying infrastructure for exchange, and ensuring that each entity fulfills its roles and responsibilities in the system. Using a common language that promotes exchangeability and interoperability across the system is probably required. Furthermore, accommodating the diverse needs of end-users is crucial for effective product utilization. Different (end-)users may have different needs when interpreting and applying the product. This thesis provided a blueprint as a framework to speed up and improve complex processes in the pathway from evidence generation to

implementation (**Chapter two**).

The subsequent chapters introduce new theoretical frameworks and tools to support complex processes in a SHE, such as evaluating and sharing clinical best practices (**Chapter three**), continuously maintaining systematic reviews and clinical practice guidelines (**Chapter four**) with priority-setting under resource constraints (**Chapters five** and **six**), bridging entities in the system to reuse each other's efforts (**Chapter seven**), and translating evidence to the end-user in understandable formats (**Chapter eight**).

Together, the thesis envisions and provides an initial evidence base for symbiotic efforts in future healthcare ecosystems. However, moving towards such a system is complex and requires a step-by-step approach rather than immediately imposing an entire system upon entities and organizations. The biggest challenges may arise from the implications on funding structures and scholarly publication models in transitioning towards a SHE. Currently, the reporting and sharing of (research) data and products by dissemination through publishers' journals are paid for by organizations and groups (e.g., article processing cost, subscriptions). Therefore, organizations might be reluctant to share the data and products they had invested in developing, despite the evident advantages of a SHE relying on unhindered and efficient data exchange. It would be interesting to observe the evolution of processes and structures to foster a culture for selfless exchange of data and products within a SHE. Nevertheless, one potential careful first step might be to start with setting-up processes that fulfill (future) roles and responsibilities of individual organizations within the context of a connected healthcare network. It is crucial to recognize the architectural perspectives and challenges associated with data compartmentalization, exchange, and the development of infrastructure to ensure the reusability and trustworthiness of products and processes throughout the network.

LITERATURE LIST

1. Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright 2013 by the National Academy of Sciences. All rights reserved.; 2013.
2. de Bruin J, Bos C, Struijs JN, Drewes HW, Baan CA. Conceptualizing learning health systems: A mapping review. Learn Health Syst. 2023;7(1):e10311.
3. Easterling D, Perry AC, Woodside R, Patel T, Gesell SB. Clarifying the concept of a learning health system for healthcare delivery organizations: Implications from a qualitative analysis of the scientific literature. Learn Health Syst. 2022;6(2):e10287.
4. Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26(2):w107-18.
5. Institute of Medicine. The Learning Healthcare System: Workshop Summary. In: Olsen L, Aisner D, McGinnis JM, editors. Washington (DC): National Academies Press (US) Copyright © 2007, National Academy of Sciences.; 2007.
6. Vandvik PO, Brandt L. Future of Evidence Ecosystem Series: Evidence ecosystems and learning health systems: why bother? J Clin Epidemiol. 2020;123:166-70.
7. Cartabellotta A, Tilson JK. The ecosystem of evidence cannot thrive without efficiency of knowledge generation, synthesis, and translation. J Clin Epidemiol. 2019;110:90-5.
8. Shepherd JP. How to achieve more effective services: the evidence ecosystem. 2014.
9. Stewart R, Dayal H, Langer L, Van Rooyen C. The evidence ecosystem in South Africa: growing resilience and institutionalisation of evidence use. Palgrave Communications. 2019;5(1).
10. Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. Ann Intern Med. 2020;173(12):1015-7.
11. Bothwell LE, Podolsky SH. The Emergence of the Randomized, Controlled Trial. N Engl J Med. 2016;375(6):501-4.
12. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Bmj. 2010;340:c332.
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372:n71.
14. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane; 2023. Available from: www.training.cochrane.org/handbook
15. Kahlert J, Gribsholt SB, Gammelager H, Dekkers OM, Luta G. Control of confounding in the analysis phase - an overview for clinicians. Clin Epidemiol. 2017;9:195-204.
16. Lee A. The development of network meta-analysis. Journal of the Royal Society of

Medicine. 2022;115(8):313-21.
17. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
18. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. Jama. 2022;328(24):2446-7.
19. 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. J Clin Epidemiol. 2020;123:153-61.
20. Nadarevic T, Colli A, Giljaca V, Fraquelli M, Casazza G, Manzotti C, et al. Magnetic resonance imaging for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. Cochrane Database Syst Rev. 2022;5(5):Cd014798.
21. Nadarevic T, Giljaca V, Colli A, Fraquelli M, Casazza G, Miletic D, Štimac D. Computed tomography for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. Cochrane Database Syst Rev. 2021;10(10):Cd013362.
22. Fischer MA, Raptis DA, Donati OF, Hunziker R, Schade E, Sotiropoulos GC, et al. MR imaging features for improved diagnosis of hepatocellular carcinoma in the non-cirrhotic liver: Multi-center evaluation. Eur J Radiol. 2015;84(10):1879-87.
23. Kim SE, Lee HC, Shim JH, Park HJ, Kim KM, Kim PN, et al. Noninvasive diagnostic criteria for hepatocellular carcinoma in hepatic masses >2 cm in a hepatitis B virus-endemic area. Liver Int. 2011;31(10):1468-76.
24. Lin MT, Wang CC, Cheng YF, Eng HL, Yen YH, Tsai MC, et al. Comprehensive Comparison of Multiple-Detector Computed Tomography and Dynamic Magnetic Resonance Imaging in the Diagnosis of Hepatocellular Carcinoma with Varying Degrees of Fibrosis. PLoS One. 2016;11(11):e0166157.
25. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. Cmaj. 2010;182(18):E839-42.
26. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. Bmj. 2016;353:i2089.
27. Alper BS, Flynn A, Bray BE, Conte ML, Eldredge C, Gold S, et al. Categorizing metadata to help mobilize computable biomedical knowledge. Learn Health Syst. 2022;6(1):e10271.
28. Alper BS, Tristan M, Ramirez-Morera A, Vreugdenhil MM, Van Zuuren EJ, Fedorowicz Z. RAPADAPTE for rapid guideline development: high-quality clinical guidelines can be rapidly developed with limited resources. Int J Qual Health Care. 2016;28(3):268-74.
29. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L, et al. Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. Int J Qual Health Care. 2006;18(3):167-76.
30. Health Level Seven International. FHIR Overview - Clinicians. 2023. Available from: https://www.hl7.org/fhir/overview-clinical.html
31. Alper BS, Dehnbostel J, Afzal M, Subbian V, Soares A, Kunnamo I, et al. Making science computable: Developing code systems for statistics, study design, and risk of bias. J Biomed Inform. 2021;115:103685.

32. Alper BS. EBMonFHIR-based tools and initiatives to support clinical research. J Am Med Inform Assoc. 2022;30(1):206-7.
33. Lichtner G, Alper BS, Jurth C, Spies C, Boeker M, Meerpohl JJ, von Dincklage F. Representation of evidence-based clinical practice guideline recommendations on FHIR. J Biomed Inform. 2023;139:104305.
34. Lichtner G, Spies C, Jurth C, Bienert T, Mueller A, Kumpf O, et al. Automated Monitoring of Adherence to Evidenced-Based Clinical Guideline Recommendations: Design and Implementation Study. J Med Internet Res. 2023;25:e41177.
35. The Cochrane Collaboration. Review Manager Web (RevMan Web). 2024.
36. Foundation MEE. MAGIC: A Digital Authoring and Publication Platform for the Evidence Ecosystem. 2024. Available from: https://app.magicapp.org/
37. McMaster University an Evidence Prime. GRADEpro GDT: GRADEpro Guideline Development Tool. 2024. Available from: https://www.gradepro.org/
38. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med. 2011;104(12):510-20.
39. Scottish Intercollegiate Guidelines Network. A guideline developer's handbook (SIGN publication no. 50). Edinburgh: SIGN; 2019. Available from: http://www.sign.ac.uk
40. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. 2023. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction
41. Schünemann H, Brożek J, Guyatt G, Oxmann A. GRADE Handbook. 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html [Accessed on: 9 August 2023]
42. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane; 2022. Available from: www.training.cochrane.org/handbook
43. Deeks JJ, Bossuyt PM, Leeflang M, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy version 2.0 (updated July 2023). Cochrane; 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current
44. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021;130:13-22.
45. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.
46. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
47. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.

48. EQUATOR Network. Enhancing the QUAlity and Transparency Of health Research. 2023. Available from: https://www.equator-network.org/ [Accessed on: 9 August 2023]
49. COMET Initiative. COMET Initiative: Core Outcome Measures in Effectiveness Trials. 2023. Available from: https://www.comet-initiative.org/, Accessed on 25 July 2023
50. Becker M, Neugebauer EA, Eikermann M. Partial updating of clinical practice guidelines often makes more sense than full updating: a systematic review on methods and the development of an updating procedure. J Clin Epidemiol. 2014;67(1):33-45.
51. Newberry SJ, Ahmadzai N, Motala A, Tsertsvadze A, Maglione M, Ansari MT, et al. AHRQ Methods for Effective Health Care. Surveillance and Identification of Signals for Updating Systematic Reviews: Implementation and Early Experience. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
52. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? Bmj. 2001;323(7305):155-7.
53. Shekelle PG, Motala A, Johnsen B, Newberry SJ. Assessment of a method to detect signals for updating systematic reviews. Syst Rev. 2014;3:13.
54. Shekelle PG, Newberry SJ, Wu H, Suttorp M, Motala A, Lim YW, et al. AHRQ Methods for Effective Health Care. Identifying Signals for Updating Systematic Reviews: A Comparison of Two Methods. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.
55. Mickenautsch S, Yengopal V. The modified Ottawa method to establish the update need of a systematic review: glass-ionomer versus resin sealants for caries prevention. J Appl Oral Sci. 2013;21(5):482-9.
56. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. J Clin Epidemiol. 2017;91:47-53.
57. El Mikati IK, Khabsa J, Harb T, Khamis M, Agarwal A, Pardo-Hernandez H, et al. A Framework for the Development of Living Practice Guidelines in Health Care. Ann Intern Med. 2022;175(8):1154-60.
58. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30.
59. Smith WD. Reweighted range voting – new multiwinner voting method. 2005.
60. Akl EA, Fadlallah R, Ghandour L, Kdouh O, Langlois E, Lavis JN, et al. The SPARK Tool to prioritise questions for systematic reviews in health policy and systems research: development and initial validation. Health Res Policy Syst. 2017;15(1):77.
61. Sanabria AJ, Alonso-Coello P, McFarlane E, Niño de Guzman E, Roqué M, Martínez García L. The UpPriority tool supported prioritization processes for updating clinical guideline questions. J Clin Epidemiol. 2021;139:149-59.
62. Sanabria AJ, Pardo-Hernandez H, Ballesteros M, Canelo-Aybar C, McFarlane E, Niño de Guzman E, et al. The UpPriority tool was developed to guide the prioritization of clinical guideline questions for updating. J Clin Epidemiol. 2020;126:80-92.
63. Kok J, Smith WD. Re-weigted Range Voting - a Proportional Representation voting

method that deels like range voting. Available from: https://www.rangevoting.org/RRV.html
64. Wiercioch W, Nieuwlaat R, Zhang Y, Alonso-Coello P, Dahm P, Iorio A, et al. New methods facilitated the process of prioritizing questions and health outcomes in guideline development. J Clin Epidemiol. 2022;143:91-104.
65. Zaror C, Deana NF, Espinoza-Espinoza G, Aravena-Rivas Y, Muñoz-Millán P, Pineda P, et al. Questions and health outcomes prioritization for the development of a COVID-19 dental clinical practice guideline: A case study. J Eval Clin Pract. 2022;28(3):404-10.

APPENDICES

Summary
Samenvatting
Dankwoord
About the author

SUMMARY

In Chapter two, the evidence ecosystem and learning healthcare system are converged into a single model to understand inefficiencies and speed-up and improve processes in the pathway from evidence generation to implementation and learning spanning both systems. This framework can be used as a blueprint to map new or existing processes and thereby identifying relevant stakeholders, pathways, infrastructure, data, and products. The chapter furthermore discusses important aspects for connecting both systems in such a symbiotic healthcare ecosystem, including the need to adjust internal organizational processes to meet responsibilities in the symbiotic system. The subsequent Chapters three to eight provide theoretical frameworks and tools to support processes for such responsibilities across a symbiotic system.

Chapter three provides a framework and a tool to evaluate healthcare provision's best practices on four axes; are context, processes, outcomes, and learning capacity. Recurring themes across all four axes seem to be culture, resources, and learning. The accompanying tool aims to capture important information about the best-practice based on indicators in the framework for evaluation purposes. Once captured, other practices can use this information for mirroring purposes in a learning health care system.

For evidence synthesizers and guideline developers becomes important to set-up maintenance strategies as they remain responsible for the state of their products. **Chapter four** provides a theoretical framework to design maintenance strategies and tailoring them to the needs and context of the organization. The chapter translates concepts of a clinical diagnostic test-treatment strategy to the context of organizational portfolio maintenance. The framework lets designers consider specific elements and tailor them to their needs to test whether there is outdatedness and which management option is indicated. The resulting maintenance strategies aim to maintain whole portfolios of the organization's systematic reviews or clinical practice guidelines.

In practice, limited resources will prevent updating every systematic review or clinical practice guideline in the organization's portfolio. Therefore, organizations need to use priority-setting assessments to select products that have the most priority to enter a maintenance strategy and/or receive specific management actions (e.g. updating). **Chapter five** extends the framework described in **Chapter four** with methodological insights about adding priority-setting assessments to maintenance strategies. Concepts from diagnostic test accuracy methodology are transferred to the context of priority-setting in maintenance strategies in this chapter. The role and outcome level of priority-setting assessments dictate their performance in a maintenance strategy, which is important to understand to efficiently utilize the available resources for maintenance.

Ultimately, organizations need to perform real-world priority-setting assessments in their maintenance strategies. **Chapter six** provides a frame of reference to help understand and structure such priority-setting assessments. Three main components are identified: a process component (any procedural step), a function component (specific rules/calculations

that determine the priority), and an outcome component (final outcome of the priority-setting assessment). Together, process and function components can be mixed and matched to meet organizational needs and reach a final priority-setting outcome with the outcome component. **Chapter six** furthermore provides a function component as a tool to aid organizations in achieving a ranked list of items. The tool is based on the re-weighted range voting mechanism and allows for a proportional representation. However, an adjusted mechanism also allows for a more disproportional representation with the potential to boost less prominent perspectives. This is important when there are important perspectives from a less represented group during the priority-setting assessment. An organization may develop procedures and use the tool according to their own needs and context without having any procedures imposed (except for meeting the tool's input format: one score per participant per item).

If the priority-setting assessment determined which guidelines had enough priority for receiving a management action (e.g., updating), there is a need to efficiently perform such actions. Here, reusing each other's efforts allows for more efficient organizational procedures. However, data and products may need to be complete and transparent, and data need to be available and accessible. **Chapter seven** provides a real-world example of bridging the gap between the evidence synthesis entity and the evidence translation entity during guideline development. The chapter illustrates how efforts in producing a complete and transparent product with enough available and accessible data in the evidence syntheses entity results in a more efficient process when developing a clinical practice guideline segment.

Developing trustworthy products, maintaining them under resource constraints, and reusing existing efforts may all be in vain when the products cannot be applied by the end-user. Therefore, **Chapter eight** connects the evidence translation community with end-users in clinical practice in the context of translating diagnostic test accuracy parameters to a more interpretable format. Using natural frequencies in a hypothetical cohort makes accuracy parameters (e.g. sensitivity, specificity) more interpretable for clinicians and patients in clinical practice, resulting in less complicated information to inform clinical decisions regarding diagnostic tests. This chapter provides considerations and guidance for selecting pretest-probabilities to translate accuracy parameters with natural frequencies.

Together, chapters in this thesis provide theoretical frameworks and tools to support processes of organizations aiming to meet specific responsibilities in a (future) symbiotic system that converges the evidence ecosystem and the learning healthcare system. Moving towards a symbiotic healthcare ecosystem is complex and probably requires a step-by-step approach. Organizations could start with setting-up processes that fulfill such (future) responsibilities within the context of a connected network, which could develop into a symbiotic system. Challenges regarding data compartmentalization, data exchange, and infrastructures for reusable and trustworthy products and processes throughout the network should be acknowledged in this development process.

SAMENVATTING

In **Hoofdstuk twee** worden het ecosysteem voor bewijs en het lerende gezondheidszorgsysteem geconvergeerd in één model om inefficiënties te begrijpen en het pad van het genereren van bewijs naar implementatie en leren te versnellen en te verbeteren, welke over beide systemen strekt. Dit raamwerk kan als blauwdruk gebruikt worden om nieuwe of bestaande processen in kaart te brengen en daarmee relevante stakeholders, paden, infrastructuur, data en producten te identificeren. Dit hoofdstuk bespreekt tevens belangrijke aspecten voor het verbinden van beide subsystemen als één symbiotisch gezondheidszorgecosysteem, waaronder de noodzaak om interne organisatorische processen aan te passen om verantwoordelijkheden in het symbiotische systeem te vervullen. De hierop volgende **Hoofdstukken drie** tot en met **acht** geven theoretische raamwerken en hulpmiddelen om processen voor dergelijke verantwoordelijkheden in het symbiotische systeem te ondersteunen.

Hoofdstuk drie biedt een raamwerk en een hulpmiddel om de beste praktijkvoeringen in de gezondheidszorg te evalueren op vier assen; context, processen, uitkomsten en leervermogen. Terugkerende thema's over de vier assen lijken cultuur, middelen en leren te zijn. Het bijgaande hulpmiddel probeert om belangrijke informatie over de beste praktijkvoering vast te leggen op basis van indicatoren uit het raamwerk voor evaluatie. Wanneer dit is vastgelegd kunnen andere praktijken deze informatie gebruiken als spiegelinformatie in een lerend gezondheidssysteem.

Voor mensen die kennis synthetiseren en voor richtlijnontwikkelaars wordt het belangrijk om onderhoudsstrategieën op te zetten, omdat zij verantwoordelijk blijven voor de staat van hun producten. **Hoofdstuk vier** biedt een theoretisch raamwerk om onderhoudsstrategieën te ontwerpen en aan te passen aan de behoeften en context van de organisatie. Het hoofdstuk vertaalt concepten van klinische diagnostische test-behandelstrategieën naar de context van on organisatorisch portfolio-onderhoud. Het raamwerk laat ontwerpers over specifieke onderdelen nadenken en deze aan hun behoeften aanpassen om te kunnen testen of er sprake is van veroudering en welke handelingen zijn geïndiceerd. De resulterende onderhoudsstrategieën zijn bedoeld om gehele portfolio's met systematische literatuuronderzoeken of richtlijnen van organisaties te kunnen onderhouden.

Vaak zullen beperkte middelen voorkomen dat elk systematisch literatuuronderzoek of richtlijn in de portfolio van de organisatie geüpdatet kan worden in de praktijk. Daarom dienen organisaties prioriteringsbeoordelingen te gebruiken om te producten te selecteren die de meeste prioriteit hebben om onderhouden te worden en/of een specifieke handeling dient te ondergaan (bijv. updaten). **Hoofstuk vijf** breidt het beschreven raamwerk uit **Hoofdstuk vier** uit met methodologische inzichten over het toevoegen van prioriteringsbeoordelingen aan onderhoudsstrategieën. Concepten uit de diagnostische testaccuratesse methodologie worden in dit hoofdstuk vertaald naar de context van prioriteringsbeoordelingen in onderhoudsstrategieën. De rol en het uitkomstniveau van prioriteringsbeoordelingen dicteren haar prestaties in een onderhoudsstrategie, wat belangrijk is om te begrijpen om de beschikbare middelen voor onderhoud efficiënt te gebruiken.

Uiteindelijk moeten organisaties daadwerkelijk prioriteringsbeoordelingen uitvoeren in hun onderhoudsstrategieën. **Hoofdstuk zes** biedt een kader om prioriteringsbeoordelingen te kunnen begrijpen en te structureren. Er zijn drie componenten geïdentificeerd: een procescomponent (elke procedurele stap), een functiecomponent (specifieke regels/ berekeningen die de prioriteit bepalen) en een uitkomstcomponent (de uiteindelijke uitkomst van de prioriteitsbeoordeling). Proces- en functiecomponenten kunnen samen worden afgestemd om aan de behoeften van de organisatie te voldoen die voor een uiteindelijke uitkomst van de prioriteringsbeoordeling zorgen in de uitkomstcomponent. **Hoofdstuk zes** biedt daarnaast ook een functiecomponent als hulpmiddel aan om organisaties te ondersteunen met het maken van een gerangschikte lijst van items. Dit hulpmiddel is gebaseerd op het zogenoemde 're-weighted range voting' mechanisme en maakt proportionele representatie mogelijk. Een aangepast mechanisme stelt het hulpmiddel ook in staat om een meer disproportionele representatie mogelijk te maken met de potentie om minder prominente perspectieven te versterken. Dit kan belangrijk zijn wanneer er belangrijke perspectieven zijn van minder vertegenwoordigde groepen tijdens de prioriteitsbeoordeling. Hierbij is het belangrijk dat de organisatie procedures kan ontwikkelen en het hulpmiddel kan gebruiken volgens de eigen behoeften en context zonder dat er enige procedures worden opgelegd (behalve het voldoen aan het invoerformat: één score per deelnemer per item).

Zodra de prioriteitsbeoordeling heeft bepaald welke, bijvoorbeeld, richtlijnen genoeg prioriteit hadden om onderhoudsacties te ontvangen (bijv. updaten), is er een noodzaak om deze acties efficiënt uit te voeren. Het hergebruik van elkaars inspanningen leidt tot efficiëntere organisatorische processen. De data en producten dienen echter compleet en transparant te zijn, met beschikbare en toegankelijke data. **Hoofdstuk zeven** biedt een voorbeeld uit de echte wereld waarin de entiteit voor het synthetiseren van bewijs wordt overbrugd naar de entiteit voor het vertalen van bewijs met richtlijnontwikkeling. Het hoofdstuk illustreert hoe inspanningen tijdens het produceren van een compleet en transparant product met genoeg beschikbare en toegankelijke data in de entiteit voor de synthese van bewijs resulteren in een efficiënter proces bij het ontwikkelen van een richtlijnmodule.

Het ontwikkelen van betrouwbare producten, het onderhouden ervan met beperkte middelen en het hergebruiken van bestaande inspanningen kunnen allemaal tevergeefs zijn wanneer het product niet kan worden toegepast door de eindgebruiker. Daarom verbindt **Hoofdstuk acht** de entiteit voor het vertalen van bewijs met de eindgebruiker in de klinische praktijk vanuit de context van het vertalen van diagnostische testaccuratesse parameters naar een begrijpelijker format. Door natuurlijke frequenties in een hypothetisch cohort te gebruiken worden accuratesseparameters (bijv. sensitiviteit en specificiteit) begrijpelijker voor clinici en patiënten in de praktijk, wat voor minder gecompliceerde informatie leidt om beslissingen over diagnostische tests te informeren. Het hoofdstuk geeft overwegingen en richting voor het selecteren van voorafkansen om accuratesseparameters te vertalen met natuurlijke frequenties. Deze vertaling zou enige indirectheid kunnen bevatten, maar dit kan acceptabel zijn gegeven de context.

Gezamenlijk bieden de hoofdstukken in dit proefschrift theoretische raamwerken en hulpmiddelen ter ondersteuning van processen van organisaties die streven naar het vervullen van specifieke verantwoordelijkheden in een (toekomstig) symbiotisch systeem

die het ecosysteem voor bewijs en het lerend gezondheidszorgsysteem convergeert. De ontwikkeling naar een symbiotisch gezondheidszorgecosysteem is complex en vereist een stapsgewijze aanpak. Organisaties zouden kunnen beginnen met het opzetten van processen die dergelijke (toekomstige) verantwoordelijkheden vervullen binnen een netwerk, welke kan uitgroeien tot een symbiotisch systeem. Uitdagingen met betrekking tot data-compartmentalisatie, het uitwisselen van data en de infrastructuren voor herbruikbare en betrouwbare producten en processen in het gehele netwerk zouden moeten worden erkend in dit groeiproces.

DANKWOORD

Dit traject had ik niet kunnen voltooien zonder ten minste de volgende personen de bedanken voor hun rol, deelname, medewerking, en/of steun in dit proces:

Prof. dr. Hooft, Lotty, bedankt voor de ruimte die je mij geboden hebt in het promotietraject qua richting en gedachten, maar bovenal voor de ruimte die je gaf om mijzelf te kunnen ontwikkelen. Niet alleen als onderzoeker, maar ook als persoon. Terugkijkend heb je als geen ander gezien wat er nodig was om dit traject tot een succes te maken. Ik wil je bedanken voor zowel de verdiepende als overstijgende discussies die ervoor gezorgd hebben dat ik telkens met waardevolle, nieuwe inzichten naar huis ging. Soms terug naar de tekentafel, maar altijd een stap in de juiste richting voor mij.

Dr. Van der Laan, Maarten, bedankt voor jouw kritische en klinische blik. Dit waren voor mij essentiële aanvullingen voor de methodologische en vaak theoretische onderwerpen waar we ons mee bezig hebben gehouden. Het zorgde ervoor dat we op het snijvlak van methoden en praktijk zijn blijven zitten (*Voor wie doen we dit allemaal?*). Jouw aanpak zorgde voor de concreetheid die ik hierin nodig had, afgewisseld met humor. Ik wil je daarom ook bedanken voor de inspirerende gesprekken en discussies in de vertaling van ideeën en plannen naar de concrete stappen en producten in dit proces.

Prof. dr. Rob Scholten, alweer een flink aantal jaren geleden ontmoetten wij elkaar voor het eerst op de Domus Medica en tot aan jouw pensioen was je mijn promotor. Ik wil je bedanken voor wijze waarop je mij vanaf het begin hebt betrokken en onderdeel van Cochrane Nederland hebt gemaakt. Hier heb ik veel aan gehad en veel van geleerd.

Graag wil ik de leden van de beoordelingscommissie, prof. dr. R. de Bree, prof. dr. H. Gardarsdottir, prof. dr. M.L. Bots, prof. dr. D.M.J. Delnoij, prof. dr. P.I. Spuls, hartelijk en met veel genoegen bedanken voor hun tijd en bereidheid om dit proefschrift te beoordelen.

Teus en Margreet, zonder de facilitatie en steun vanuit het Kennisinstituut was het natuurlijk nooit gelukt. Ik waardeer ik in het bijzonder de ruimte die mij werd geboden naast de reguliere werkzaamheden gedurende het traject. Dank voor de vrijheid om zelf onderwerpen te kiezen die tevens van belang zijn voor de praktijk van richtlijnontwikkeling, ik heb dit als zeer prettig ervaren. Naast de werkzaamheden waren er natuurlijk ook de borrels, diners, feesten, quizzen, cursussen en (buitenlandse) congressenbezoeken die mogelijk werden gemaakt door het Kennisinstituut. Ook hiervoor veel dank, een aanzienlijk deel van het plezier gedurende het traject is hieraan te danken (met blijvende herinneringen aan Santiago de Chile).

Roy en Annefloor, ondanks dat we niet samen het hele traject hebben doorlopen heb ik van jullie geleerd, waarvoor veel dank. Roy, ook enorm bedankt voor jouw input, kritische blik en aanhoudende inzet als coauteur van twee hoofdstukken gedurende de afgelopen jaren. Annefloor, bedankt voor jouw optimisme en enthousiasme bij aanvang van het traject.

Fijne collegae van het Kennisinstituut, dank voor de heerlijke werksfeer en de gezelligheid naast al het harde werken. Marlies en Ester, jullie zijn na mij begonnen met een PhD-traject, maar hebben het eerder afgerond. Betere voorbeelden had ik daar niet in kunnen hebben. Bedankt voor alle fijne gesprekken over werk, onderzoek en de balans. Charlotte en Femke, ontzettend leuk om Kennisinstituut collegae als coauteurs op een aantal hoofdstukken te hebben. Bedankt voor jullie enthousiasme, samengaand met de vlotte en fijne samenwerking. Ook Team Onco (tegenwoordig 1 en 2) wil ik bedanken voor hun steun en geduld als ik weer eens over 'het ecosysteem' begon, maar ook voor de leuke momenten op en buiten het kantoor. De wintersport-groepen van 2019 en 2020 (incl summer escape): bedankt voor de gezelligheid en dat ik jullie ook op deze wijze heb mogen leren kennen. Skiën kan ik niet, maar dat maakte jullie niets uit.

Fijne collegae van Cochrane Nederland, heel fijn hoe jullie mij op hebben genomen in het team ondanks dat ik niet altijd frequent aanwezig kon zijn op het Julius. René, met een glimlach denk ik terug aan alle discussies wij hebben voortgebracht in het Julius Centrum. Dit was niet alleen leuk, maar het verbreedde mijn horizon en kijk op de (wetenschappelijke) wereld. Laten we dit doorzetten. Ik kan je daarnaast niet volledig bedanken zonder ook te zeggen dat ik jou als mentor erg heb gewaardeerd tijdens het traject. Soms advies over werk-inhoudelijke zaken en soms over persoonlijke dingen, dank je wel. Anneke, Pauline, Robin en Kevin, bedankt voor alle goede, inspirerende en gezellige gesprekken, steun en medeleven, koffietjes en lunches én natuurlijk de prettige samenwerking. Kevin, ook bedankt voor de fijne samenwerking op één van de hoofdstukken. Kim, Demy, Mike, Linde en Emilie het is altijd leuk om op het Julius te komen dankzij jullie! Heel fijn dat we samen aan mooie projecten en onderwijs kunnen werken. Julia, dank voor de goede gesprekken die altijd voor nieuwe inspiratie en inzichten zorgen, ongeacht waar we het over hebben.

Hans, ookal zien we elkaar niet zo vaak meer op de woensdagen, dank voor jouw interesse en het onuitputtelijke en aanstekelijke enthousiasme.

Tom, erg leuk hoe we samenwerken in de werkplaats. Ik ga altijd met plezier naar onze overleggen, bedankt!

Charlotte, bedankt voor de fijne samenwerking vanuit Groningen en voor het vergroten van mijn begrip over best-practices.

Alle coauteurs op de hoofdstukken, dank voor jullie enthousiasme, inzet en input op de betreffende stukken. Dit zijn mijns inziens stukken geworden waar we trots op mogen zijn, dank.

Oud-kamergenoten in het Julius Centrum, met plezier denk ik terug aan de pre-corona-tijd waar we met z'n allen in één kantoorruimte zaten. Het was een gezellige tijd waarin we samen hard konden werken, maar ook hard konden lachen. Dank jullie wel.

Lieve schoonfamilie, ik voel mij ontzettend welkom in jullie midden. Bedankt voor de leuke verjaardagen, uitjes, en etentjes die voor de prettige afleiding tijdens het traject zorgden.

Rik, van de sociëteit van Tabula Rasa en Yogyakarta naar gazamenlijke lunches en over de kids praten, het leven verandert. We zien elkaar niet vaak, maar het zit altijd goed. Bedankt!

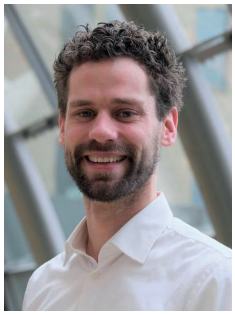
Peter, Peter, Daan, Anna, Geert-Jan, Krystel, bedankt voor jullie vriendschap en warmte. We komen al jaren bij elkaar over de vloer en doen leuke dingen samen. Van verjaardagen en (baby)bezoekjes tot spelletjes(dagen) en weekendjes weg. Dit waardeer ik enorm en het zorgde bij mij voor het nodige plezier en afleiding tijdens het traject, dank jullie wel!

Lieve familie, Jan en Luci, Lotte en Jurre met Vik en Eli, Moniek en Tim, mijn 'studie' zit erop. Bedankt dat jullie altijd klaar staan en ik bij jullie terecht kan. Jullie vertrouwde en liefdevolle aanwezigheid is altijd erg fijn. Lotte en Moniek, heel fijn dat jullie mijn paranimfen willen zijn en dit met mij willen delen. Moniek, ook bedankt voor jouw hulp en kritische blik tijdens het ontwerpen van het proefschrift.

Lieve Esther, zonder jou zou het proces voor mij veel zwaarder zijn geweest. Jouw vrolijke noot is van onschatbare waarde voor mij en niet alleen in het promotietraject, maar ook in de gang van ons leven. Tijdens alle leuke en minder leuke veranderingen en 'life events' in de afgelopen jaren ben je er voor mij geweest. Es, samen staan we sterk en ik ben gek op jou.

Lieve Tom, lieve zoon, bedankt voor de trots, de vreugde en het plezier die je mij nu al geeft in jouw jonge leven. We gaan er samen iets moois van maken.

ABOUT THE AUTHOR



Michiel Sebastiaan Oerbekke was born in Almelo, the Netherlands, on the 8th of September, 1988. He completed his secondary education in 2005 at the r.k. scholengemeenschap St. -Canisius in Almelo. After a first attempt in higher professional education (Art & Technology, Saxion University of Applied Sciences, Enschede) and working for a short while, he re-started his higher professional education with the Bachelor in Physical Therapy in 2007 (including a minor in musculoskeletal physical therapy), which he obtained in 2011 at the Saxion University of Applied Sciences. In the subsequent period from 2011 up to 2015, he worked part-time as a physical therapist in various first line practices. In this period, Michiel also started and completed his adult pre-university education in math (ROC Enschede, Enschede, 2011-2012), started and completed his premaster in Clinical Health Sciences (Utrecht University, Utrecht, 2012-2013), and eventually graduated cum laude from the master Clinical Health Sciences (Utrecht University, Utrecht, 2013-2015) to obtain his master's degree.

Michiel then became a teacher/researcher at Saxion University of Applied Sciences in Enschede in the Bachelor in Podiatry. He supervised bachelor theses and clinical case studies, coordinated education in the bachelor's final year, wrote program-wide examination procedures, and provided practical skills-based training regarding the basic clinical examination of the lumbar spinal cord, pelvis, and lower extremities.

In 2017 he joined the Knowledge Institute of the Dutch Association Medical Specialists (Utrecht), where he started as a junior consultant and was provided with the opportunity to simultaneously start a PhD-education in collaboration with Cochrane Netherlands. Up until present, Michiel advises clinical practice guideline panels about guideline methodology as a consultant and performs evidence synthesis for these guidelines, usually in the field of oncology. Together with his colleagues at the Knowledge Institute, he provided both the 'Evidence Based Guideline Development' course for guideline panel members and in-house seminars for new employees of the Knowledge Institute, covering topics such as risk of bias, evidence-to-decision, and diagnosis over the years.

At Cochrane Netherlands, located in in the Julius Center for Health Sciences and Primary Care (Utrecht), he simultaneously started as a PhD-candidate in 2017. Michiel's appreciation for providing higher education continued during his PhD-trajectory. One of the more notable educational activities was designing and providing two-day courses in rapid evidence synthesis tailored to the specific needs of units from the National Institute for Public Health and the Environment (RIVM) during the COVID-19 pandemic, together with his respective colleagues at Cochrane Netherlands.

In 2023 he accepted a part-time secondment to Cochrane Netherlands to provide daily supervision in projects informing policies concerning the impact of digital healthcare on the well-being of healthcare professionals. in 2024 this secondment was extended, where he

now also contributes to commissioned evidence syntheses for clinical practice guidelines of the American College of Physicians at Cochrane Netherlands. Simultaneously, at the Kennisinstituut, Michiel will increase his focus to translate and implement the knowledge and insights obtained through the research in this thesis.

NOTABLE ACTIVITIES

Courses

Systematic reviews of diagnostic studies
Advanced diagnostic research
Clinical trials and drug risk assessment
Public health epidemiology
Clinical epidemiology
Systematic reviews in intervention research
Evidence-based guideline development

Master classes

Becoming the editor's pick: key strategies for drafting manuscripts for publication (Amsterdam, the Netherlands)

Presentations

Goed gebruik geneesmiddelen, oral (Den Bosch, the Netherlands)
Cochrane Colloquium, poster (Edinburgh, Scotland, UK)
MEMTAB, poster (Utrecht, the Netherlands)

(Inter)national conferences

Goed gebruik geneesmiddelen (Den Bosch, the Netherlands)
Cochrane Colloquium (Edinburgh, Scotland, UK)
MEMTAB (Utrecht, the Netherlands)

Teaching / supervision

ROBINS-I risk of bias practical (master epidemiology, Utrecht, the Netherlands)
Introduction to meta-analyses (master epidemiology, Utrecht, the Netherlands)
Supervision of post-master researchers (Utrecht/Groningen, the Netherlands
Evidence based practice for Basalt Revalidatie (Den Haag, the Netherlands)
Searching efficiently and reading critically (University college, Utrecht, the Netherlands)
Supervising master student's thesis (master biomedical science, Radboud University, Nijmegen, the Netherlands)
Rapid review for the National Institute for Public Health and the Environment (online, Cochrane Netherlands, Utrecht)
Evidence-based guideline development (Kennisinstituut, Utrecht, the Netherlands)
STARTblok P3, evidence based case report (master GNK, Utrecht, the Netherlands)
Searching efficiently and reading critically (bachelor GNK, Utrecht, the Netherlands)
Guidelines and GRADE for paramedics (Utrecht, the Netherlands)

Scientific publications not in this thesis

Kok-Pigge AC, Greving JP, de Groot JF, **Oerbekke M**, Kuijpers T, Burgers JS. A Delphi consensus checklist helped assess the need to develop rapid guideline recommendations. J Clin Epidemiol. 2023 Apr;156:1-10. doi: 10.1016/j.jclinepi.2023.02.007. Epub 2023 Feb 9. PMID: 36764465.

Jenniskens K, Bootsma MCJ, Damen JAAG, **Oerbekke MS**, Vernooij RWM, Spijker R, Moons KGM, Kretzschmar MEE, Hooft L. Effectiveness of contact tracing apps for SARS-CoV-2: a rapid systematic review. BMJ Open. 2021 Jul 12;11(7):e050519. doi: 10.1136/bmjopen-2021-050519. PMID: 34253676; PMCID: PMC8277487.

Lankhorst K, **Oerbekke M**, van den Berg-Emons R, Takken T, de Groot J. Instruments Measuring Physical Activity in Individuals Who Use a Wheelchair: A Systematic Review of Measurement Properties. Arch Phys Med Rehabil. 2020 Mar;101(3):535-552. doi: 10.1016/j.apmr.2019.09.006. Epub 2019 Oct 10. PMID: 31606452.

Oerbekke MS, Stukstette MJ, Schütte K, de Bie RA, Pisters MF, Vanwanseele B. Concurrent validity and reliability of wireless instrumented insoles measuring postural balance and temporal gait parameters. Gait Posture. 2017 Jan;51:116-124. doi: 10.1016/j.gaitpost.2016.10.005. Epub 2016 Oct 6. PMID: 27744250.

Evidence syntheses for clinical practice guidelines (published and in development)

Development of diagnostic, clinimetric, and/or intervention modules in:Richtlijn amputatie en prothesiologie onderste extremiteit, Richtlijn colorectaal carcinoom, Richtlijn fysieke fitheid bij oncologische patiënten, Richtlijn galweg- en galblaascarcinoom, Richtlijn hepatocellulair carcinoom, Richtlijn hoofd-halstumoren, Richtlijn liesbreuk bij kinderen, Richtlijn pancreascarcinoom,