AI in evidence synthesis and guidelines: Should we trust the machines?

Zachary Munn

October 2027

THE UNIVERSITY of ADELAIDE

1

2

We acknowledge and pay our respects to the Kaurna people, the traditional custodians whose ancestral lands we gather on.

We acknowledge the deep feelings of attachment and relationship of the Kaurna people to country and we respect and value their past, present and ongoing connection to the land and cultural beliefs.

THE UNIVERSITY of ADELAIDE

3

Interests/Disclosures/Overview

JBI ADELAIDE

Director HESRI, School of Public Health

GRADE CENTRE

GRADE Centre Director and Working Group Member

GIN Guidelines International Network

Past-Chair Guidelines International Network (GIN)

Funding for Systematic Reviews and Guidelines

JBI

Head of Evidence Synthesis Taxonomy Initiative

PERSyst Prevalence Estimates Reviews

Involved in many Methodology Groups/Networks

NHMRC Investigator Grant Funding APP1195676

THE UNIVERSITY of ADELAIDE

Website: https://health.adelaide.edu.au/hesri/

4

METHODOLOGY ARTICLE

The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI)

Zachary Munn,[1] Edoardo Aromataris,[1] Catalin Tufanaru,[1] Cindy Stern,[1] Kylie Porritt,[1] James Farrow,[2] Craig Lockwood,[1] Matthew Stephenson,[1] Sandeep Moola,[1] Lucylynn Lizarondo,[1] Alexandra McArthur,[1] Micah Peters,[1] Alan Pearson[1] and Zoe Jordan[1]

[1]The Joanna Briggs Institute, The University of Adelaide, Adelaide, South Australia, [2]Farrow Norris, Sydney, Australia

ABSTRACT

Aim: Systematic reviews play an important role in ensuring trustworthy recommendations in healthcare. However, systematic reviews can be laborious to undertake and as such software has been developed to assist in the conduct and reporting of systematic reviews. The Joanna Briggs Institute and its collaborating centres consist of thousands of researchers, academics and clinicians across the globe conducting systematic reviews of various types. To support them in their work, modern software and online tools are required. Our aim was to develop a software program to support systematic reviewers across the globe.

Methods: A working party was formed with extensive consultation with members of the Joanna Briggs Collaboration focusing on ideal features of a software program to support systematic reviews. The new systematic review software was built using an agile methodology and designed to be a modern web application.

Results: The new systematic review software, the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI), was successfully developed through an iterative process of development, feedback, testing and review. The software is now available (https://www.jbisumari.org/) and supports the entire systematic review process for different types of systematic reviews.

Conclusions: An agile software development approach combined with wide consultation and user testing can facilitate systematic review software design and development. This new software can support systematic reviews and guideline developers to create systematic reviews for a diverse range of questions.

Key words: Joanna Briggs Institute, meta-analysis, software, systematic review

Int J Evid Based Healthc 2019; 17:36–43.

Disclaimer:

Product developer (once upon a time)

Not an AI or ML expert

THE UNIVERSITY of ADELAIDE

5

Interests/Disclosures/Overview

….I use AI all the time*

Our team use AI to assist in guideline development (with checks and oversight)

- Creation of health outcome descriptors

- SWAR of AI supported searching, screening and selection

- Support integrity checking

- Double check RoB assessments

- Plain language conversions

- Wording of survey preambles

- Check R code

+++++

***note: no AI was used in the development of this presentation (except where credited) – I am accountable for all mistakes/errors and terrible formatting in this presentation**

THE UNIVERSITY of ADELAIDE

6

Challenges in Synthesis and Guidelines

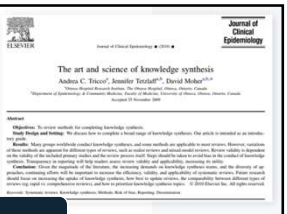
THE UNIVERSITY of ADELAIDE

7

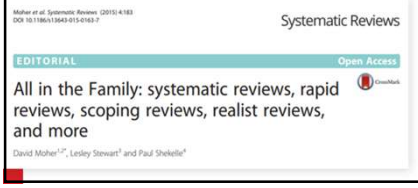
The Evidence Synthesis Family

Journal of Clinical Epidemiology
Volume 73, May 2016, Pages 19-28

A scoping review identifies multiple er
synthesis methods, but few studies op


The art and science of knowledge synthesis

- Systematic reviews
- Mixed methods reviews
- Qualitative evidence synthesis
- Rapid reviews
- Realist synthesis
- Umbrella/Overviews Reviews
- Evidence (gap) maps
- Scoping reviews


Systematic Reviews

EDITORIAL Open Access

All in the Family: systematic reviews, rapid reviews, scoping reviews, realist reviews, and more

David Moher[1,2*], Lesley Stewart[3] and Paul Shekelle[4]

….and more!

Clarifying differences between review designs
d methods

David Gough, James Thomas and Sandy Oliver

Systematic Reviews 2012, 1:28

https://doi.org/10.1186/2046-4053-1-28 | © Gough et al.; licensee BioMed Central Ltd. 2012

Received: 4 January 2012 | Accepted: 19 March 2012 | Published: 9 June 2012

Open Peer Review reports

8

Types of Systematic Reviews

1. **Interventions/Effectiveness**
2. **Qualitative/Experiences**
3. **Costs/Economic Evaluations**
4. **Prevalence or Incidence**
5. **Diagnostic Test Accuracy**
6. **Etiology and Risk**
7. **Policy/Opinion**
8. **Measurement properties**
9. **Prognostic**
10. **Methodological**
11. **Preferences and values**
12. **Animal/bench studies**

…and more!

Munn et al. BMC Medical Research Methodology (2018) 18:5
DOI 10.1186/s12874-017-0468-4

BMC Medical Research Methodology

CORRESPONDENCE Open Access

What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences

Zachary Munn*, Cindy Stern, Edoardo Aromataris, Craig Lockwood and Zoe Jordan

Abstract

Background: Systematic reviews have been considered as the pillar on which evidence-based healthcare rests. Systematic review methodology has evolved and been modified over the years to accommodate the range of questions that may arise in the health and medical sciences. This paper explores a concept still rarely considered by novice authors and in the literature: determining the type of systematic review to undertake based on a research question or priority.

Results: Within the framework of the evidence-based healthcare paradigm, defining the question and type of systematic review to conduct is a pivotal first step that will guide the rest of the process and has the potential to impact on other aspects of the evidence-based healthcare cycle (evidence generation, transfer and implementation). It is something that novice reviewers (and others not familiar with the range of review types available) need to take account of but frequently overlook. Our aim is to provide a typology of review types and describe key elements that need to be addressed during question development for each type.

Conclusions: In this paper a typology is proposed of various systematic review methodologies. The review types are defined and situated with regard to establishing corresponding questions and inclusion criteria. The ultimate objective is to provide clarified guidance for both novice and experienced reviewers and a unified typology with respect to review types.

Keywords: Systematic reviews, Evidence-based healthcare, Question development

THE UNIVERSITY of ADELAIDE

9

Task	Description	Classification
1. formulate review question	Decide on the research question of the review.	preparation
2. find previous SR	Search for SR that answers the same question.	
3. write the protocol	Provide an objective, reproducible, sound methodology for peer review.	
4. devise search strategy	Decide on databases and keywords to find all relevant trials.	
5. search	Aim to find all relevant citations even if many irrelevant ones included.	retrieval
6. de-duplicate	Remove identical citations.	
7. screen abstracts	Based on titles and abstracts, remove definitely-irrelevant trials.	appraisal
8. obtain full text	Download, request copies from authors, inter-library loans, etc.	
9. screen full text	Exclude irrelevant trials.	
10. snowball	Follow citations from included trials to find additional trials.	retrieval
11. extract data	Extract outcome numbers and associate with trial arm.	synthesis
12. synthesize data	Convert extracted data to common representation (usually average and SD).	
13. re-check literature	Repeat the search to find new literature published since the initial search.	retrieval
14. meta analyze	Statistically combine the results from all included trials.	synthesis
15. write up review	Produce and publish the final report.	write-up

Systematic review automation technologies

Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani & Enrico Coiera

Systematic Reviews 3, Article number: 74 (2014) | Cite this article

THE UNIVERSITY of ADELAIDE

10

Issues and Challenges in Evidence Synthesis

1. Approximately 80 reviews a day
2. Duplication/Redundant reviews
3. Poor quality/poorly reported
4. No impact in practice
5. Costly and time consuming
6. Out of date by the time they are completed
7. Incorrect approach/methods for the question

(Hoffman 2021, Møller 2018, Ioannidis 2016, Page 2016, Munn 2018)

THE UNIVERSITY of ADELAIDE

11

Ensuring high quality and efficient development of trustworthy guidelines

Planning and production - GIN McMaster checklist

Handbooks and guidance – NHMRC, GRADE

Reporting standards – RIGHT

Evaluate trustworthiness of a guideline – AGREE

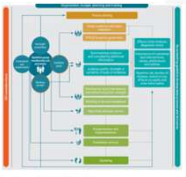
Software – GRADEPro, MAGICApp, Covidence

Networking, registering guidelines, collaboration - GIN

Adapt and adopt – GRADE Adolopment

Engagement with community – MuSE/ GIN Public

For education and training – INGUIDE



GRADE

RIGHT
Reporting Items for Practice Guidelines in Healthcare

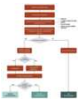
AGREE II

Authoring & Publication Platform

covidence

GIN
Guidelines International Network

GIN Public



MuSE

INGUIDE
International Guideline Development Credentialing & Certification Program

THE UNIVERSITY of ADELAIDE

12

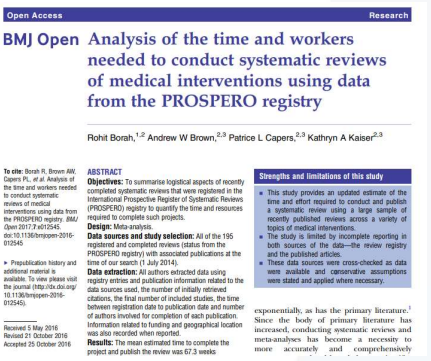
Guidelines - Costly and timely

Systematic reviews average 67.3 weeks
(Borah 2017)

Average guideline takes 18-30 months
(NHMRC)

Costs up to approximately 1 million AUD, USD $775000, 5 million CNY (NHMRC)

Countless person hours



Open Access Research

BMJ Open

Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry

Rohit Borah,[1,2] Andrew W Brown,[2,3] Patrice L Capers,[2,3] Kathryn A Kaiser[2,3]

To cite: Borah R, Brown AW, Capers PL, *et al*. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 2017;7:e012545. doi:10.1136/bmjopen-2016-012545

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2016-012545).

Received 5 May 2016
Revised 21 October 2016
Accepted 25 October 2016

ABSTRACT

Objectives: To summarise logistical aspects of recently completed systematic reviews that were registered in the International Prospective Register of Systematic Reviews (PROSPERO) registry to quantify the time and resources required to complete such projects.

Design: Meta-analysis.

Data sources and study selection: All of the 195 registered and completed reviews (status from the PROSPERO registry) with associated publications at the time of our search (1 July 2014).

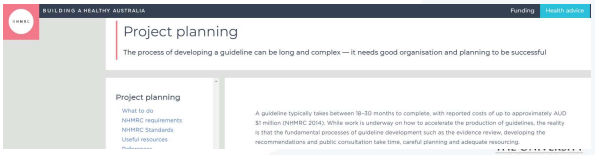
Data extraction: All authors extracted data using registry entries and publication information related to the data sources used, the number of initially retrieved citations, the final number of included studies, the time between registration date to publication date and number of authors involved for completion of each publication. Information related to funding and geographical location was also recorded when reported.

Results: The mean estimated time to complete the project and publish the review was 67.3 weeks

Strengths and limitations of this study

- This study provides an updated estimate of the time and effort required to conduct and publish a systematic review using a large sample of recently published reviews across a variety of topics of medical interventions.
- The study is limited by incomplete reporting in both sources of the data—the review registry and the published articles.
- These data sources were cross-checked as data were available and conservative assumptions were stated and applied where necessary.

exponentially, as has the primary literature.[1] Since the body of primary literature has increased, conducting systematic reviews and meta-analyses has become a necessity to more accurately and comprehensively



BUILDING A HEALTHY AUSTRALIA

Funding Health advice

Project planning

The process of developing a guideline can be long and complex — it needs good organisation and planning to be successful

Project planning

What to do
NHMRC requirements
NHMRC Standards
Useful resources

A guideline typically takes between 18–30 months to complete, with reported costs of up to approximately AUD $1 million (NHMRC 2014). While work is underway on how to accelerate the production of guidelines, the reality is that the fundamental processes of guideline development such as the evidence review, developing the recommendations and public consultation take time, careful planning and adequate resourcing.

THE UNIVERSITY of ADELAIDE

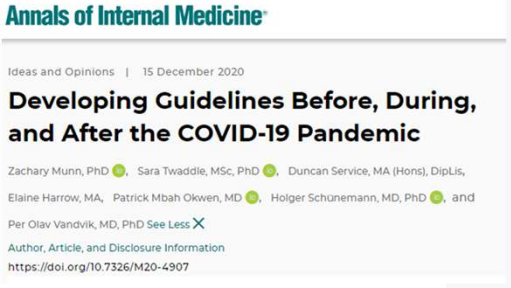
Guidelines - Inefficient

Much duplication in effort

Lack of automation

Hesitancy to adopt systematic review and guideline development tools

Lack of collaboration and sharing



Annals of Internal Medicine

Ideas and Opinions | 15 December 2020

Developing Guidelines Before, During, and After the COVID-19 Pandemic

Zachary Munn, PhD, Sara Twaddle, MSc, PhD, Duncan Service, MA (Hons), DipLis, Elaine Harrow, MA, Patrick Mbah Okwen, MD, Holger Schünemann, MD, PhD, and Per Olav Vandvik, MD, PhD See Less

Author, Article, and Disclosure Information

https://doi.org/10.7326/M20-4907

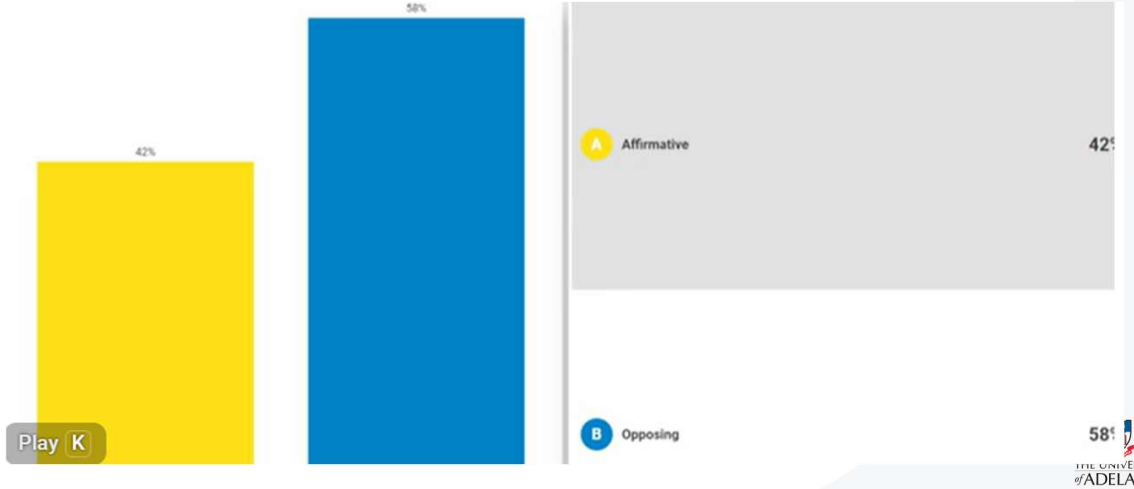
"The process of developing trustworthy recommendations has often been plagued with **inefficiencies, redundancies, duplication**, and **a lack of collaboration and sharing**. "

"Members of the guideline and evidence-based health care communities have been working for years to improve this poorly functioning evidence ecosystem, yet the underlying collaborations, tools, and processes for wholescale efficiencies in the translation of evidence into practice were not widely established when the COVID-19 pandemic began."

THE UNIVERSITY of ADELAIDE

15

Should AI replace humans in evidence synthesis?



16

Agenda

1. AI and automation – a historical perspective
2. How AI works
3. Tools and uptake
4. What can't AI replace
5. Workflows and considerations

THE UNIVERSITY of ADELAIDE

17

Timeline of Systematic Reviews

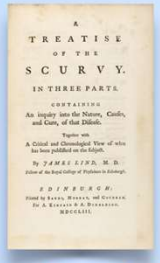
18

Timeline of Systematic Reviews

1753 — 1904 — 1976 — 1970s — 1992 — 1992 — 1995 — 1996 — 2000 — 2002 — 2010 — 2011 — 2014 — 2021

19

Timeline of Systematic Reviews


A
TREATISE
OF THE
SCURVY.
IN THREE PARTS.
CONTAINING
An inquiry into the Nature, Cause, and Cure, of that Disease.
Together with
A Critical and Chronological View of what has been published on the Subject.
By JAMES LIND, M. D.
Fellow of the Royal College of Physicians in Edinburgh.
EDINBURGH:
MDCCLIII.

Treatise of the Scurvy

1753 — 1904 — 1976 — 1970s — 1992 — 1992 — 1995 — 1996 — 2000 — 2002 — 2010 — 2011 — 2014 — 2021

20

Timeline of Systematic Reviews

Treatise of the Scurvy

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

Pearson statistically combines studies on typhoid vaccine

21

Timeline of Systematic Reviews

Treatise of the Scurvy

Glass coins the term 'meta-analysis'

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

Pearson statistically combines studies on typhoid vaccine

22

Timeline of Systematic Reviews

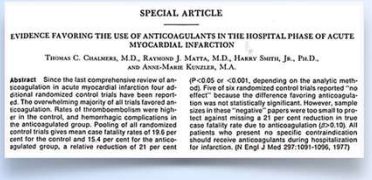
Treatise of the Scurvy

Glass coins the term 'meta-analysis'

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

Pearson statistically combines studies on typhoid vaccine

Early exemplar systematic reviews conducted



SPECIAL ARTICLE

EVIDENCE FAVORING THE USE OF ANTICOAGULANTS IN THE HOSPITAL PHASE OF ACUTE MYOCARDIAL INFARCTION

THOMAS C. CHALMERS, M.D., RAYMOND J. MATTA, M.D., HARRY SMITH, JR., PH.D., AND ANNE-MARIE KUNZLER, M.A.

Abstract Since the last comprehensive review of anticoagulation in acute myocardial infarction four additional randomized control trials have been reported. The overwhelming majority of all trials favored anticoagulation. Rates of thromboembolism were higher in the control, and hemorrhagic complications in the anticoagulated group. Pooling of all randomized control trials gives mean case fatality rates of 19.6 per cent for the control and 15.4 per cent for the anticoagulated group, a relative reduction of 21 per cent (P<0.05 or <0.001, depending on the analytic method). Five of six randomized control trials reported "no effect" because the difference favoring anticoagulation was not statistically significant. However, sample sizes in these "negative" papers were too small to protect against missing a 21 per cent reduction in true case fatality rate due to anticoagulation (β=0.10). All patients who present no specific contraindication should receive anticoagulants during hospitalization for infarction. (N Engl J Med 297:1091-1096, 1977)

23

Timeline of Systematic Reviews

Treatise of the Scurvy

Glass coins the term 'meta-analysis'

Evidence-based medicine coined by Gordon Guyatt

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

Pearson statistically combines studies on typhoid vaccine

Early exemplar systematic reviews conducted

24

Timeline of Systematic Reviews

Treatise of the Scurvy

Glass coins the term 'meta-analysis'

Evidence-based medicine coined by Gordon Guyatt

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

Pearson statistically combines studies on typhoid vaccine

Early exemplar systematic reviews conducted

Cochrane Collaboration Established

Cochrane

25

Timeline of Systematic Reviews

Treatise of the Scurvy

Glass coins the term 'meta-analysis'

Evidence-based medicine coined by Gordon Guyatt

PICO framework proposed

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

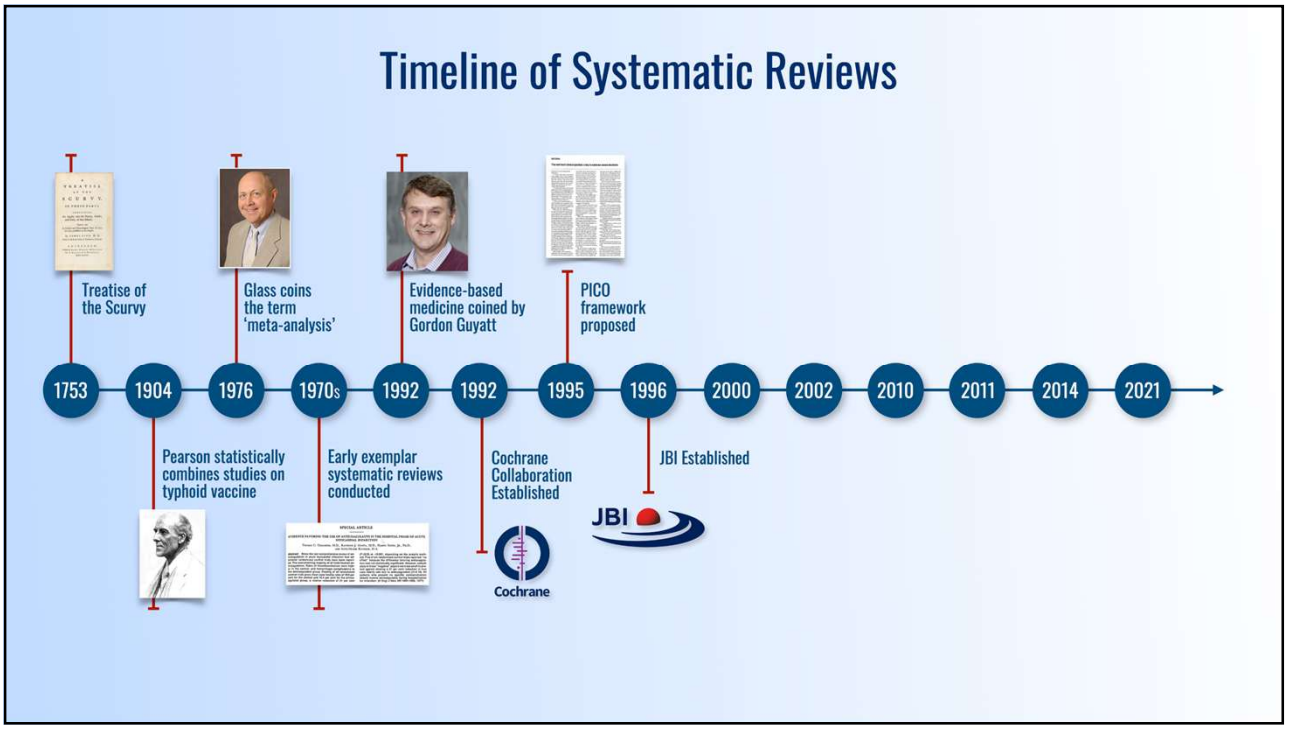
Pearson statistically combines studies on typhoid vaccine

Early exemplar systematic reviews conducted

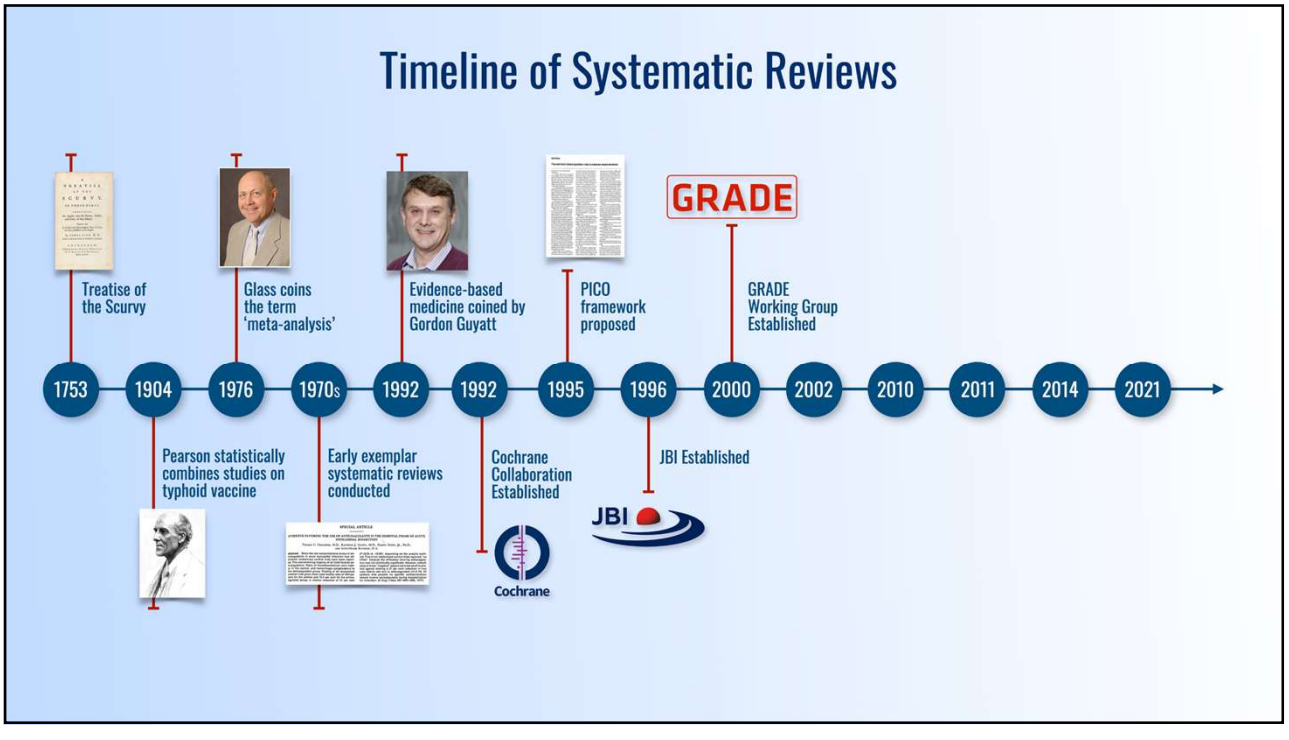
Cochrane Collaboration Established

Cochrane

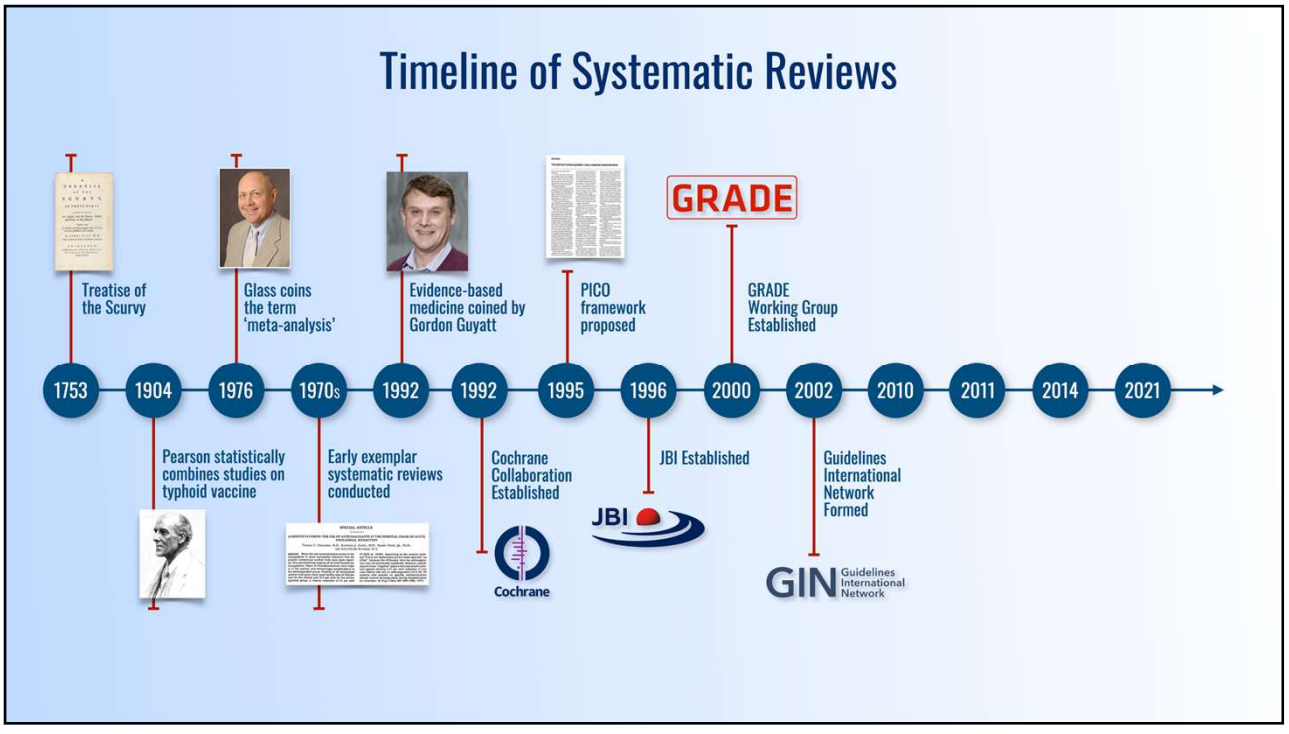
26


Timeline of Systematic Reviews
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
JBI
Cochrane

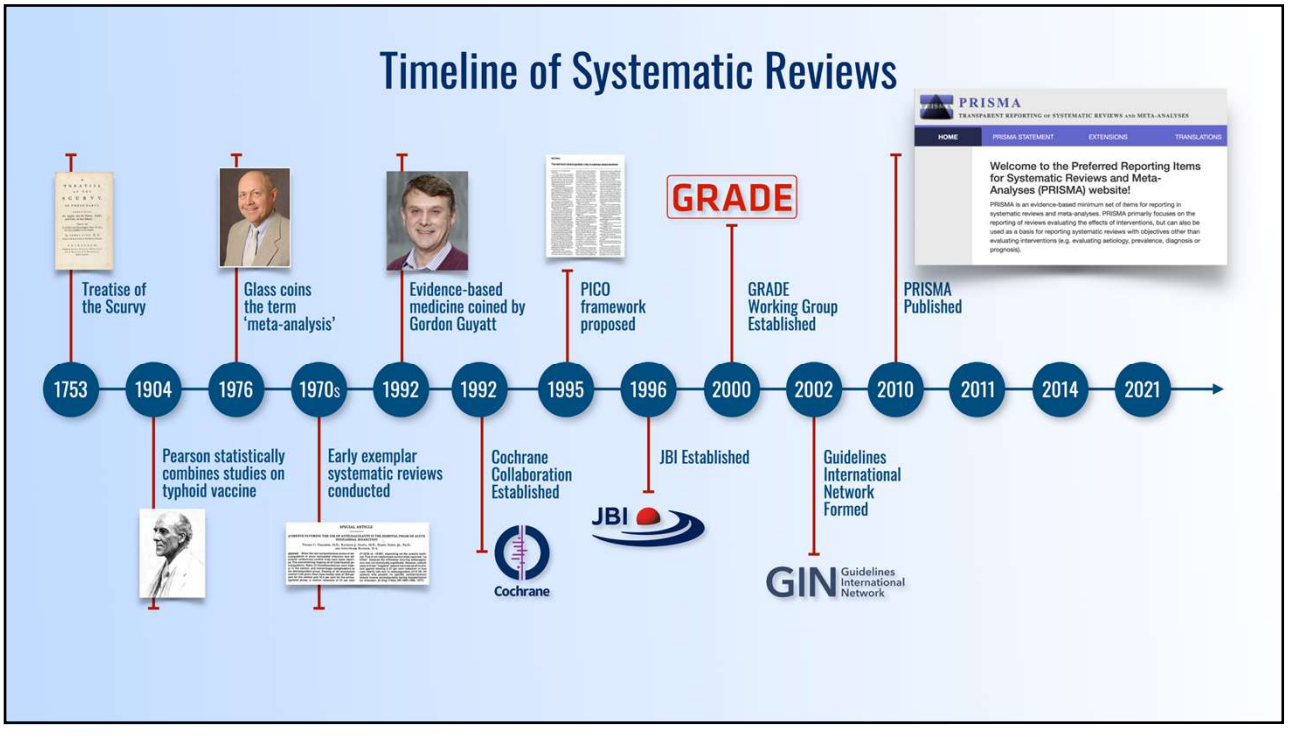
27


Timeline of Systematic Reviews
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
GRADE
GRADE Working Group Established
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
JBI
Cochrane

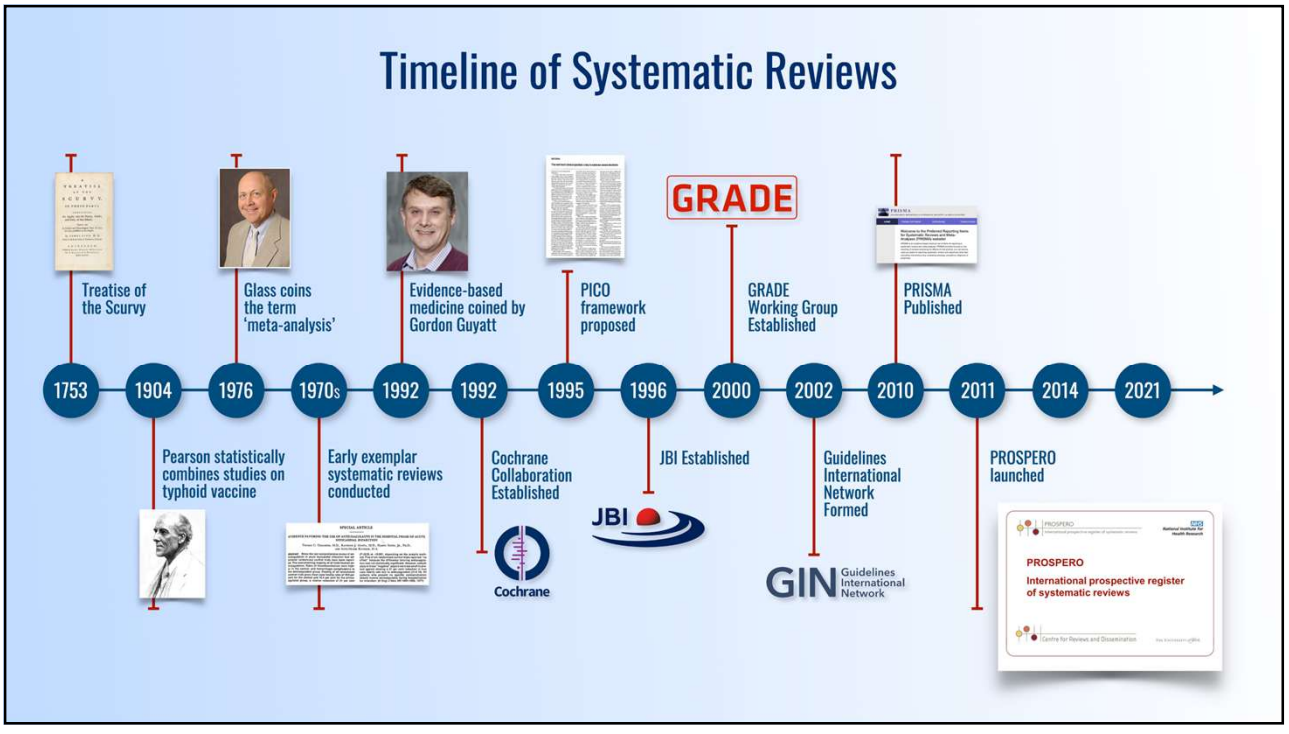
28


Timeline of Systematic Reviews
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
GRADE
GRADE Working Group Established
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
Guidelines International Network Formed
JBI
Cochrane
GIN Guidelines International Network

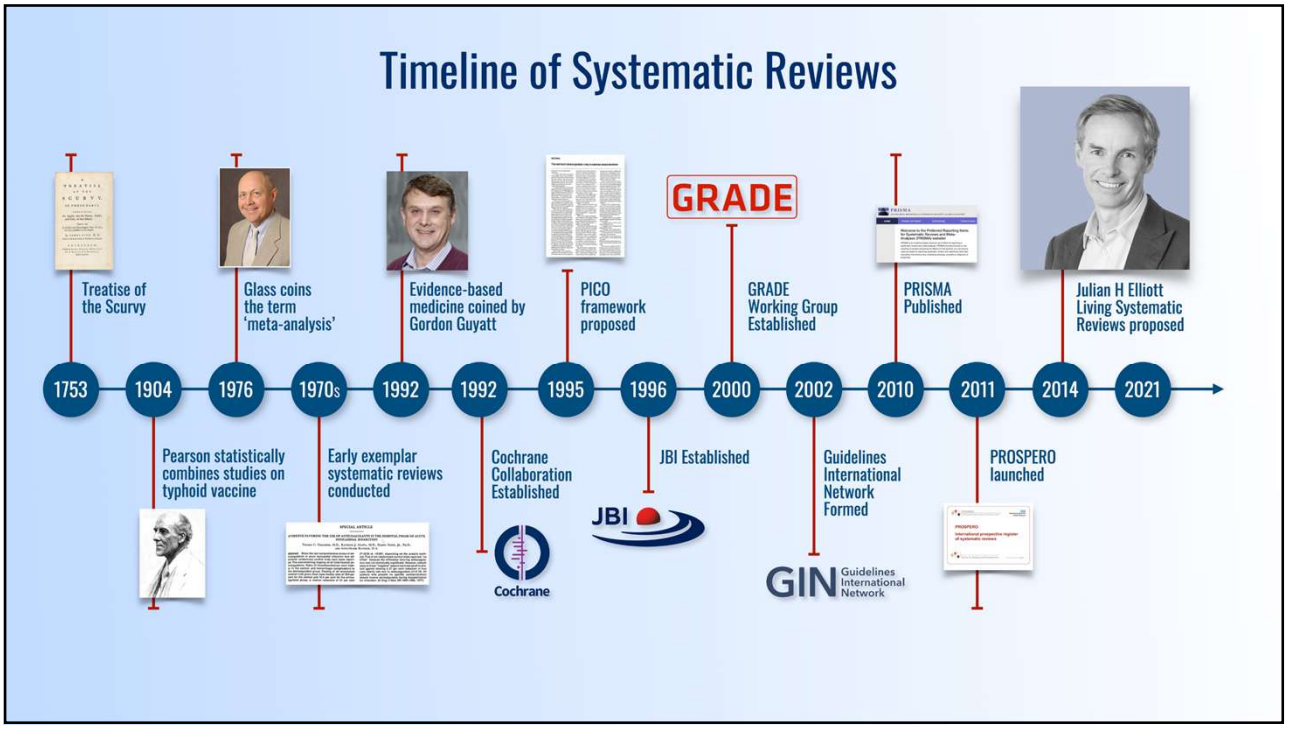
29


Timeline of Systematic Reviews
PRISMA
TRANSPARENT REPORTING OF SYSTEMATIC REVIEWS AND META-ANALYSES
HOME
PRISMA STATEMENT
EXTENSIONS
TRANSLATIONS
Welcome to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website!
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
GRADE
GRADE Working Group Established
PRISMA Published
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
Guidelines International Network Formed
JBI
Cochrane
GIN Guidelines International Network

30

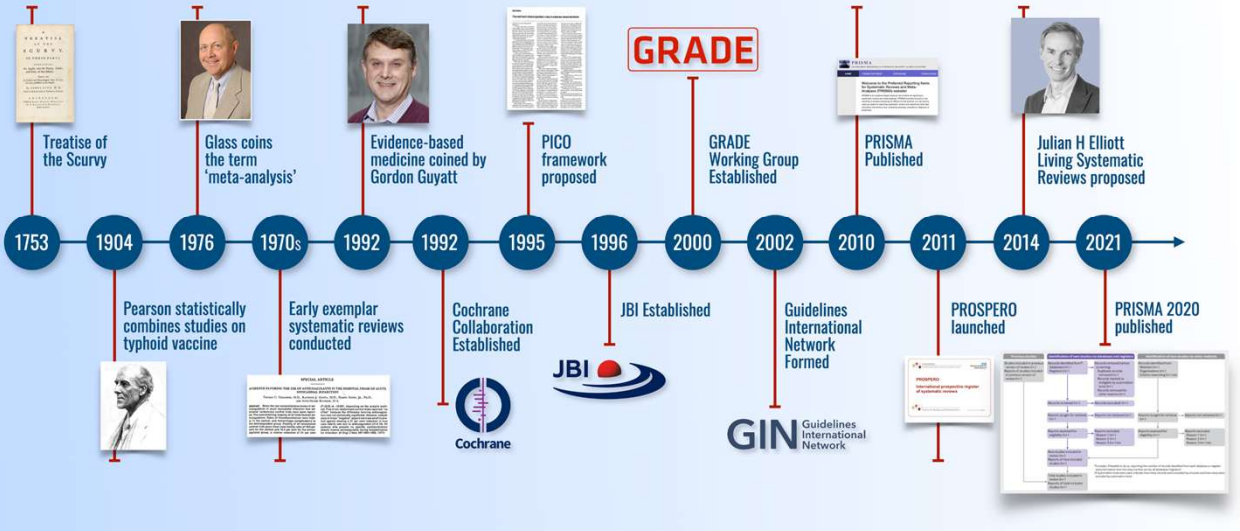

Timeline of Systematic Reviews
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
GRADE
GRADE Working Group Established
PRISMA Published
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
Guidelines International Network Formed
PROSPERO launched
Cochrane
JBI
GIN Guidelines International Network
PROSPERO
International prospective register of systematic reviews

31


Timeline of Systematic Reviews
Treatise of the Scurvy
Glass coins the term 'meta-analysis'
Evidence-based medicine coined by Gordon Guyatt
PICO framework proposed
GRADE
GRADE Working Group Established
PRISMA Published
Julian H Elliott Living Systematic Reviews proposed
1753
1904
1976
1970s
1992
1992
1995
1996
2000
2002
2010
2011
2014
2021
Pearson statistically combines studies on typhoid vaccine
Early exemplar systematic reviews conducted
Cochrane Collaboration Established
JBI Established
Guidelines International Network Formed
PROSPERO launched
Cochrane
JBI
GIN Guidelines International Network

32

Timeline of Systematic Reviews



Timeline of Systematic Reviews

Pre-tools

Early review specific tools
RevMan (1994)
SUMARI) 2002
Statistical software
Clunky but beneficial

The future…

1753 1904 1976 1970s 1992 1992 1995 1996 2000 2002 2010 2011 2014 2021

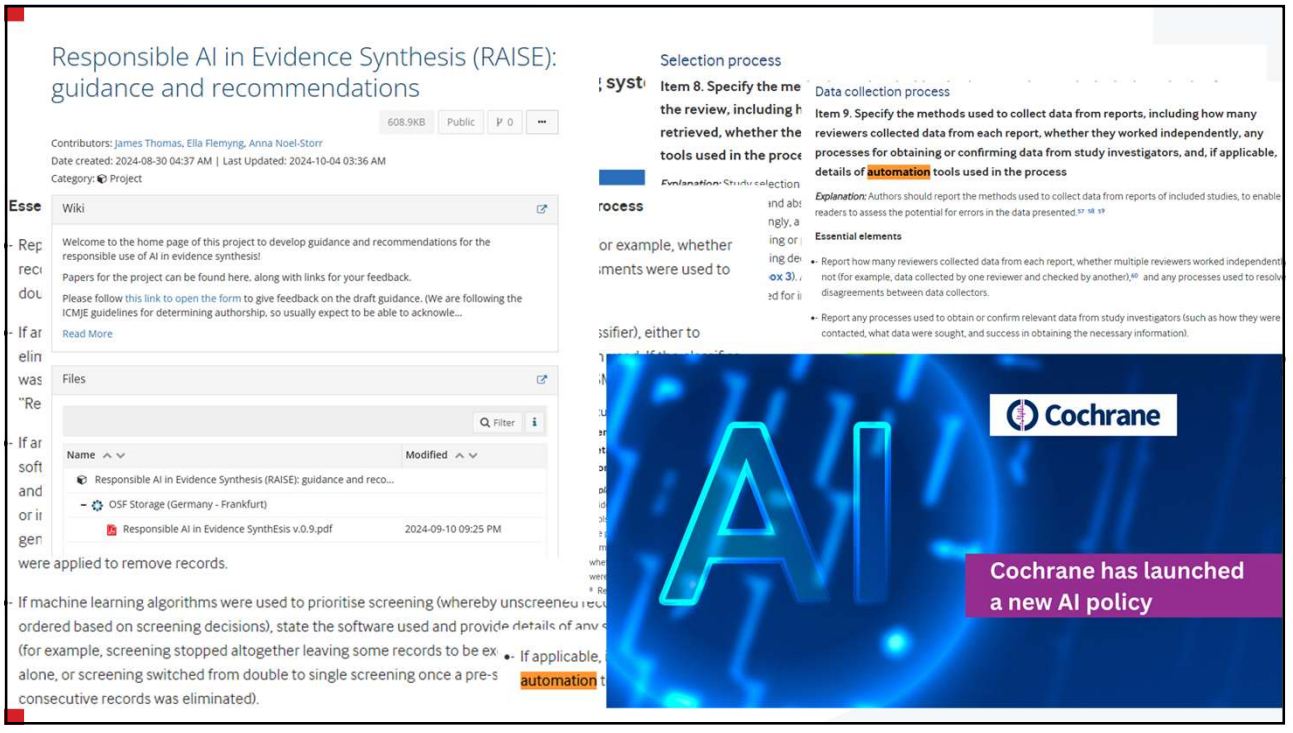
Early Tools
Reference management (80s)
Online Searching databases

Explosion in tools
SR Toolbox (2014)

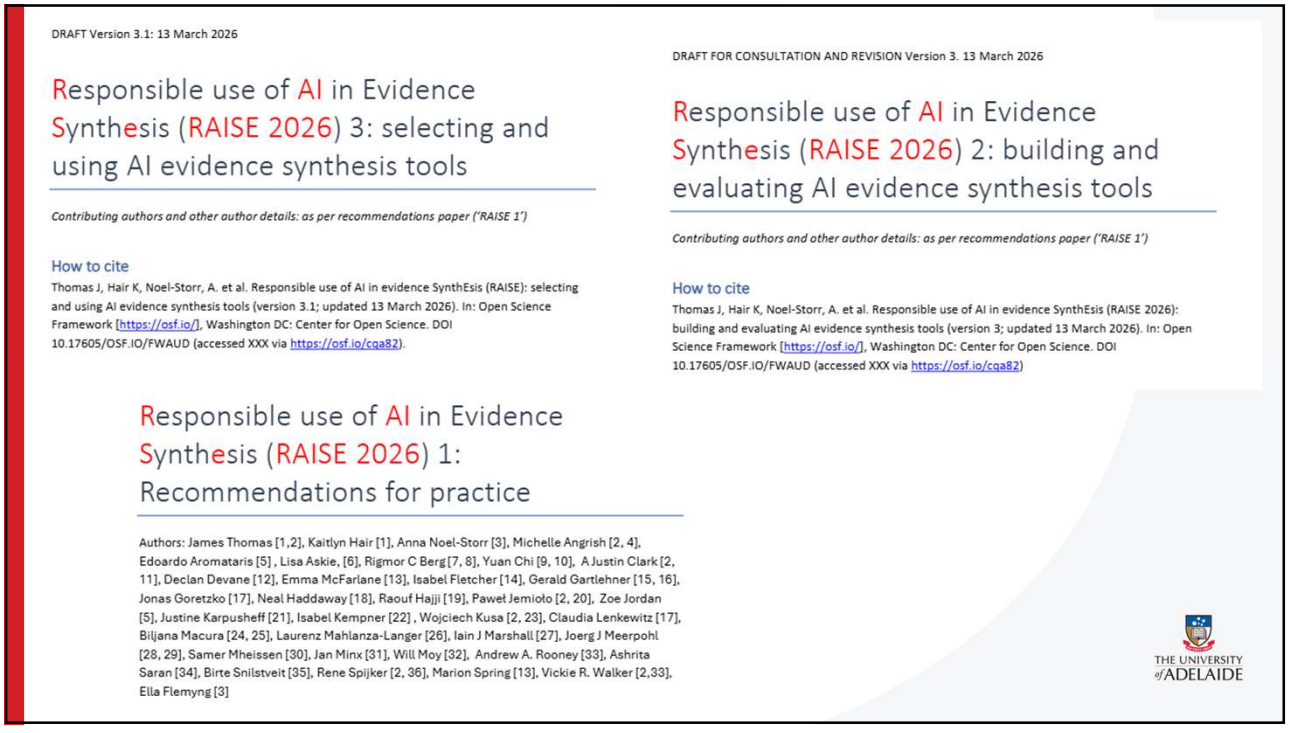
Interoperability (to an extent)
More automation
Improved features
Better UI
Mass use
Efficiency improvements

The automation of systematic reviews
Would lead to best currently available evidence at the push of a button

Guy Tsafnat *senior research fellow*[1], Adam Dunn *research fellow*[1], Paul Glasziou *professor*[2], Enrico Coiera *professor*[1]


Responsible AI in Evidence Synthesis (RAISE): guidance and recommendations
608.9KB
Public
Contributors: James Thomas, Ella Flemyng, Anna Noel-Storr
Date created: 2024-08-30 04:37 AM | Last Updated: 2024-10-04 03:36 AM
Category: Project
Wiki
Welcome to the home page of this project to develop guidance and recommendations for the responsible use of AI in evidence synthesis!
Papers for the project can be found here, along with links for your feedback.
Please follow this link to open the form to give feedback on the draft guidance. (We are following the ICMJE guidelines for determining authorship, so usually expect to be able to acknowle...
Read More
Files
Filter
Name
Modified
Responsible AI in Evidence Synthesis (RAISE): guidance and reco...
OSF Storage (Germany - Frankfurt)
Responsible AI in Evidence SynthEsis v.0.9.pdf
2024-09-10 09:25 PM
were applied to remove records.
Selection process
Item 8. Specify the me
the review, including h
retrieved, whether the
tools used in the proce
Data collection process
Item 9. Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and, if applicable, details of automation tools used in the process
Explanation: Authors should report the methods used to collect data from reports of included studies, to enable readers to assess the potential for errors in the data presented.
Essential elements
Report how many reviewers collected data from each report, whether multiple reviewers worked independently or not (for example, data collected by one reviewer and checked by another), and any processes used to resolve disagreements between data collectors.
Report any processes used to obtain or confirm relevant data from study investigators (such as how they were contacted, what data were sought, and success in obtaining the necessary information).
Cochrane
AI
Cochrane has launched a new AI policy
If machine learning algorithms were used to prioritise screening (whereby unscreened records are ordered based on screening decisions), state the software used and provide details of any settings used (for example, screening stopped altogether leaving some records to be ex
If applicable, automation
alone, or screening switched from double to single screening once a pre-s
consecutive records was eliminated).

35


DRAFT Version 3.1: 13 March 2026
Responsible use of AI in Evidence Synthesis (RAISE 2026) 3: selecting and using AI evidence synthesis tools
Contributing authors and other author details: as per recommendations paper ('RAISE 1')
How to cite
Thomas J, Hair K, Noel-Storr, A. et al. Responsible use of AI in evidence SynthEsis (RAISE): selecting and using AI evidence synthesis tools (version 3.1; updated 13 March 2026). In: Open Science Framework [https://osf.io/], Washington DC: Center for Open Science. DOI 10.17605/OSF.IO/FWAUD (accessed XXX via https://osf.io/cqa82).
DRAFT FOR CONSULTATION AND REVISION Version 3. 13 March 2026
Responsible use of AI in Evidence Synthesis (RAISE 2026) 2: building and evaluating AI evidence synthesis tools
Contributing authors and other author details: as per recommendations paper ('RAISE 1')
How to cite
Thomas J, Hair K, Noel-Storr, A. et al. Responsible use of AI in evidence SynthEsis (RAISE 2026): building and evaluating AI evidence synthesis tools (version 3; updated 13 March 2026). In: Open Science Framework [https://osf.io/], Washington DC: Center for Open Science. DOI 10.17605/OSF.IO/FWAUD (accessed XXX via https://osf.io/cqa82)
Responsible use of AI in Evidence Synthesis (RAISE 2026) 1: Recommendations for practice
Authors: James Thomas [1,2], Kaitlyn Hair [1], Anna Noel-Storr [3], Michelle Angrish [2, 4], Edoardo Aromataris [5] , Lisa Askie, [6], Rigmor C Berg [7, 8], Yuan Chi [9, 10], A Justin Clark [2, 11], Declan Devane [12], Emma McFarlane [13], Isabel Fletcher [14], Gerald Gartlehner [15, 16], Jonas Goretzko [17], Neal Haddaway [18], Raouf Hajji [19], Paweł Jemioło [2, 20], Zoe Jordan [5], Justine Karpusheff [21], Isabel Kempner [22] , Wojciech Kusa [2, 23], Claudia Lenkewitz [17], Biljana Macura [24, 25], Laurenz Mahlanza-Langer [26], Iain J Marshall [27], Joerg J Meerpohl [28, 29], Samer Mheissen [30], Jan Minx [31], Will Moy [32], Andrew A. Rooney [33], Ashrita Saran [34], Birte Snilstveit [35], Rene Spijker [2, 36], Marion Spring [13], Vickie R. Walker [2,33], Ella Flemyng [3]
THE UNIVERSITY of ADELAIDE

36

Cochrane Database of Systematic Reviews | Editorial

Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence 2025

Ella Flemyng, Anna Noel-Storr, Biljana Macura, Gerald Gartlehner, James Thomas, Joerg J. Meerpohl, Zoe Jordan, Jan Minx, Angelika Eisele-Metzger, Candyce Hamel, Paweł Jemioło, Kylie Porritt, Matthew Grainger Authors' declarations of interest

Version published: 11 November 2025

https://doi.org/10.1002/14651858.ED000178

Key messages:

1. Evidence synth
   use artificial int
2. Cochrane, the
   the aims of the
   a framework fo

synthesis ecosystem.

3. Evidence synthesists developing and p
   Collaboration for Environmental Evide
   compromise the methodological rigou
4. AI and automation in evidence synthes
5. Any use of AI or automation that makes
   synthesis report.
6. AI tool developers should proactively ensure their AI systems or tools adhere to the RAISE recommendations so we have clear, transparent and publicly available information to inform decisions about whether an AI system or tool could and should be used in evidence synthesis.

Based on RAISE, Cochrane's position statement outlines four expectations for evidence synthesists using AI:

1. You are ultimately responsible for your research, including the decision to use AI and how it is used
2. You can use AI provided you can demonstrate it will not compromise the methodological rigour or integrity of your synthesis
3. AI use should be fully and transparently reported
4. AI should be used with human oversight

Expectations 1 and 2 are where you assess an AI tool to help make an informed decision on whether to use it. But what steps do you need to take to do that?

THE UNIVERSITY of ADELAIDE

37

Table 2: The CESAR project's stopping boundaries and decisions rules for interim analyses

Performance metrics	Futility boundaries (point estimate)	Non-inferiority margins (Upper limit of 95% CI)*	Decision rules
	Screening		
Sensitivity	<80%	<95%	Stop if either boundary is crossed
Specificity (for full-text screening only)	<50%	<60%	Stop if either boundary is crossed
	Data extraction		
Sensitivity	<92%	<97%	Stop if either boundary is crossed
Major error proportion	>3%	>2%	Stop if either boundary is crossed
	Usability		
System usability scale (score)	<57	<75	Stop if threshold is not met

Right tool, right job: deciding when to not use an AI tool

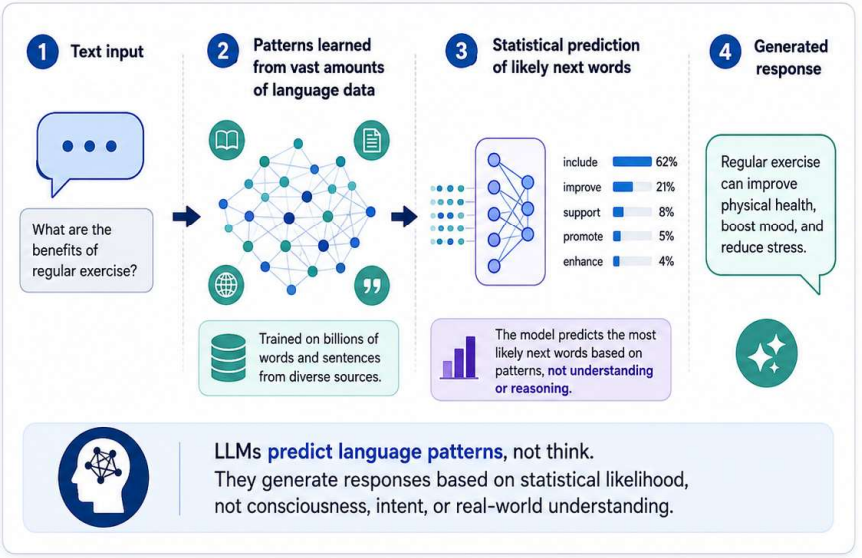
What factors should you consider when deciding whether to use AI tools to support your work?

Monday, June 15, 2026 • By Ella Flemyng, Cochrane's Head of Editorial Policy and Research Integrity

38

How LLMs Work

Generative AI (LLMs) provide a statistical representation of language — in contrast to human reasoning



Key limitations

- Rule based calculations
- Don't always like uncertainty, nuance
- Unclear decisions
- Makes up fake information / hallucinates
- Perpetuates structural bias, stereotypes, etc

**made with ChatGPT*

39

If we use AI…

Summary Table of Common Parameters

Parameter	Goal	Lower Setting Behavior	Higher Setting Behavior
Temperature	Creativity/Randomness	Factual, repetitive	Creative, risky
Top-P (p)	Diversity of word choice	Limited, safe choices	Diverse, broad vocabulary
Top-K	Limit potential words	Highly restrictive	Broad, risky
Presence	Topic variety	Repetitive topics	Introduces new topics
Frequency	Sentence structure	Exact repetition	Varied phrasing

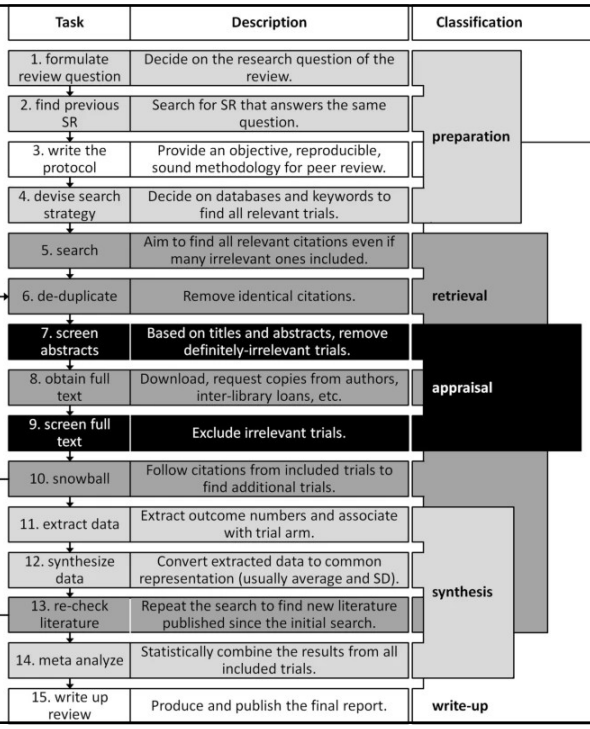
UNIVERSITY OF ADELAIDE

40

Goal	Temperature	Top-P	Penalty
Factual / Coding	Low (0.2)	Low (0.1)	Low
Standard Chat	Medium (0.7)	High (0.9)	Medium
Creative Writing	High (1.0+)	High (1.0)	High

THE UNIVERSITY of ADELAIDE

41



Systematic review automation technologies

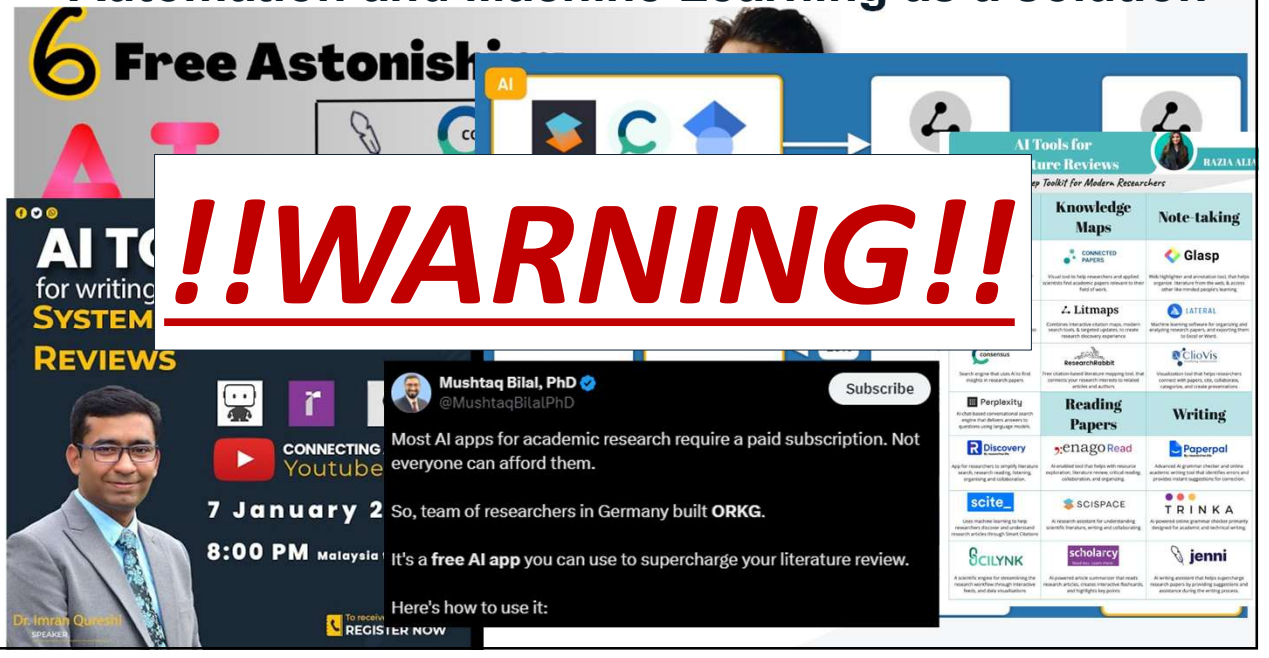
Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani & Enrico Coiera

Systematic Reviews 3, Article number: 74 (2014) | Cite this article

THE UNIVERSITY of ADELAIDE

42

Automation and Machine Learning as a solution



Automation tools to support undertaking scoping reviews

The International Collaboration for Automation of Systematic Reviews (ICASR) has put forward a set of principles for researchers to use automation for systematic reviews highlighting that automation has the potential to facilitate the production of systematic reviews that adhere to high standards for the reporting, conduct, and updating of rigorous reviews.

The Vienna Principles

1. Systematic reviews involve multiple tasks, each with different issues, but all must be improved.
2. Automation may assist with all tasks, from scoping reviews to identifying research gaps as well protocol development to writing and dissemination of the review.
3. The processes for each task can and should be continuously improved, to be more efficient and more accurate.
4. Automation can and should facilitate the production of systematic reviews that adhere to high standards for the reporting, conduct and updating of rigorous reviews.
5. Developments should also provide for flexibility in combining and using, e.g. subdividing or merging steps and allowances for different users to use different interfaces.
6. Different groups with different expertise are working on different parts of the problem; to improve reviews as a whole will require collaboration between these groups.
7. Every automation technique should be shared, preferably by making code, evaluation data and corpora available for free.
8. All automation techniques and tools should be evaluated using a recommended and replicable method with results and data reported.

ICASR

International Collaboration for the Automation of Systematic Reviews

THE UNIVERSITY of ADELAIDE

SR tools, automation, streamlining

ELSEVIER

Journal of Clinical Epidemiology 138 (2021) 80–94

ORIGINAL ARTICLE

Systematic review automation tools improve efficiency b knowledge impedes their adoption: a survey

Anna Mae Scott[a,*], Connor Forbes[a], Justin Clark[a], Matt Carter[a], Paul Zachary Munn[b]

[a]Institute for Evidence-Based Healthcare, Bond University, Gold Coast, Australia
[b]JBI, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia

Accepted 30 June 2021; Available online 7 July 2021

Abstract

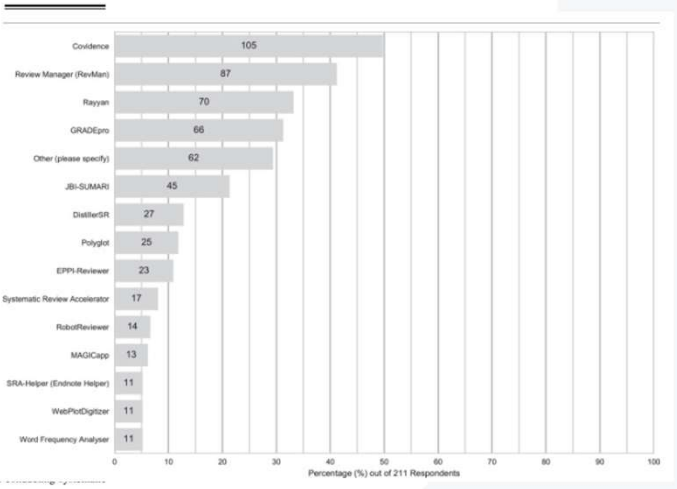
Objective: We investigated systematic review automation tool use by systematic reviewers, health technolo guideline developerst.

Study design and setting: An online, 16-question survey was distributed across several evidence synt assessment and guideline development organizations. We asked the respondents what tools they use and aband do they use the tools, their perceived time savings and accuracy, and desired new tools. Descriptive statistics results.

Results: A total of 253 respondents completed the survey; 89% have used systematic review automation whilst screening (79%). Respondents' "top 3" tools included: Covidence (45%), RevMan (35%), Rayyan and most commonly abandoned were Rayyan (19%), Covidence (15%), DistillerSR (14%) and RevMan (13%). Too increased accuracy (54%). Respondents taught themselves to how to use the tools (72%); lack of knowledg barrier to tool adoption (51%). New tool development was suggested for the searching and data extraction sta

Conclusion: Automation tools will likely have an increasingly important role in high-quality and timely required in training and dissemination of automation tools and ensuring they meet the desirable features of tho reviews. © 2021 Elsevier Inc. All rights reserved.

Keywords: Automation; Automation tools; Systematic review; Systematic review automation; Health technology assessment; Clinical practice guideline



THE UNIVERSITY of ADELAIDE

45

Table 2. Speed and accuracy improvement from automation tool use

How much time do these tools save?	
Time saved	**N (%) of 184 responses**
Saves a lot of time	67 (36%)
Saves some time	81 (44%)
Neutral	28 (15%)
Costs some time	6 (3%)
Costs lots of time	2 (1%)
Do the tools improve the accuracy of the systematic review/HTA/CPG?	
Accuracy	**N (%) of 185 responses**
A lot of improvement in accuracy	50 (27%)
A little improvement in accuracy	50 (27%)
Neutral	82 (44%)
Lose some accuracy with using the tools	3 (2%)
Lose a lot of accuracy with using the tools	0 (0%)

Abbreviations: HTA, health technology assessment; CPG, clinical practice guideline.

THE UNIVERSITY of ADELAIDE

46

Table 3. Factors preventing respondents from using the automation tools

What prevents respondents from using the automation tools	
Factors	**N (%) of 196 responses**
The learning curve (too complicated to learn)	76 (39%)
Time constraints	55 (28%)
Costs	88 (45%)
Language barrier	5 (3%)
Internet connectivity/firewall	5 (3%)
User unfriendly	49 (25%)
Lack of knowledge of existing tools	100 (51%)
Not applicable	20 (10%)
Other (optional: please share why)	43 (22%)

THE UNIVERSITY of ADELAIDE

47

METHODOLOGY PAPER

Are systematic review and guideline development tools useful? A Guidelines International Network survey of user preferences

Zachary Munn PhD, GradDip HlthSc, BMedRad,[1] Linn Brandt MD,[2] Ton Kuijpers PhD,[3] Craig Whittington PhD,[4] Louise Wiles PhD, BPhysio[5,6,7] Torsten Karge[8,9], on behalf of the G-I-N Tech Working Group

[1]Faculty of Health and Medical Sciences, JBI, The University of Adelaide, Adelaide, South Australia, Australia, [2]Department of Medicine, Innlandet Hospital Trust, Gjøvik, Norway, [3]Dutch College of General Practitioners, Utrecht, The Netherlands, [4]Doctor Evidence, Santa Monica, California, USA, [5]Australian Centre for Precision Health, School of Health Sciences, University of South Australia, Adelaide, South Australia, [6]Faculty of Medicine and Health Sciences, Centre for Healthcare Resilience and Implementation Science, Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia, [7]South Australian Health and Medical Research Institute (SAHMRI), Adelaide, South Australia, [8]Clinical Guideline Services, and [9]Medical Department I (Gastroenterology, Infectiology, and Rheumatology), Charité – Universitätsmedizin, Berlin, Germany

ABSTRACT

Background: There are now over 140 tools/programs that can assist in developing systematic reviews or clinical practice guidelines (CPGs). It is currently unclear which tools are used by systematic reviewers and CPG developers, which development processes they are used for, and what facilitators or barriers to their use exist.

Methods: To determine which tools are currently being used by systematic reviewers and CPG developers, an online survey was administered during July–August 2017. Guidelines International Network individual and organizational members were invited to participate. Survey questions focused on the nature and frequency of members' use of tools to support systematic review and CPG development.

Results: The overall response rate was 34%. The largest number of respondents developed one to five guidelines a year (48%). GRADEpro GDT was the most popular tool (26% of respondents) followed by Dropbox (16%) and RevMan (14%). From the options provided, the reason most respondents (85%) used particular tools was 'to be more efficient'. Most users stated they would use the tool again (95%), and 95% would recommend it to other organizations. However, respondents reported that tool efficiency and facilitators such as data sharing functionality were offset by their availability and cost, issues with structured data fields (that did not allow customization), and other technical and usability factors (e.g., features, workflows).

Conclusion: The results of this survey provide a focus for discussing improvements in tools to meet the needs of systematic reviewers and CPG developers, and a basis from which to test the efficacy and appropriateness of various tools and platforms across a number of purposes and contexts.

Key words: guidelines, Guidelines International Network, systematic review, technology

JBI Evid Implement 2020; 18:345–352.

Research highlights
- There are many different tools available to assist with the conduct of systematic reviews and guidelines.
- The survey identified the most commonly used tools amongst members of an international network (Guidelines International Network) and how frequently they are used.
- Barriers to using tools include the cost of some tools, lack of customization and lack of features.
- Facilitators for using tools included improving the efficiency of the process, ability to share data/information, and being free to use.

Correspondence: Zachary Munn, PhD, GradDip HlthSc, BMedRad, Faculty of Health and Medical Sciences, JBI, The University of Adelaide, Adelaide 5005, SA, Australia. Tel: +61 8 8313 3642; e-mail: Zachary.Munn@adelaide.edu.au

DOI: 10.1097/XEB.0000000000000226

Background

Systematic reviews, and the recommendations or clinical practice guidelines (CPGs) arising from them, form the basis of evidence-based healthcare. Over time, systematic review and guideline development methods have become increasingly demanding and complex. As such, the development of systematic reviews and trustworthy CPGs require significant investment of resources, with systematic reviews alone taking anywhere from 6 months to 2 years to complete,[1–3] with an average development time of 67.3 weeks shown in one study.[4] With the need to provide up-to-date and trustworthy recommendations in clinical practice, this

Tool name[a]	Users
GRADEpro	23
Dropbox	14
RevMan	12
Other tool	11[b]
Covidence	8
No online tool	8
MAGICapp	7
Rayyan	6
Google docs	6
PEER	5
DistillerSR	4
Team collaboration tools	4[c]
Clinical guideline services	2
EPPI Reviewer	2
DECIDE Tools (via epistemonikos)	1
GROWTH DOC Library/DOC Data	1
JBI SUMARI	1
Abstrackr	1
Box	1
Microsoft Sharepoint	1

[a]No respondents reported using the following tools: Can Implement, Systematic Review Data Repository, and CREBP Review assistant.
[b]Endnote 4, SurveyMonkey 2, AGREE or AGREE-II 3, Web based forum 1, SIGN Evidence Management System 1.
[c]Projectplace, Zoho, Basecamp, Customised guidelines wiki.

THE UNIVERSITY of ADELAIDE

48

For which step in the guideline development process does your organization use this tool?	Totals
Organisation, budget, planning and training	13
Priority setting	8
Guideline group membership	10
Establishing guideline group processes – including COI declarations + management processes	10
Identifying target audience and topic selection	3
Consumer and stakeholder involvement	11
Conflict of Interest considerations	9
(PICO) question generation	19
Write the protocol	11
Consider importance of outcomes and interventions, values, preferences and utilities	9
Devise search strategy (academic librarian input, test)	7
Search	14
De-duplicate	17
Screen abstracts (title and abstract)	31
Obtain full text	16
Screen full text	24
Snowball/pearl/checking reference list/citations	6
Assess risk of bias	25
Extract data	25
Re-check literature	8
Meta-analyse	19
Summarizing evidence and considering additional information	32
Judging quality, strength or certainty of a body of evidence	29
Developing recommendations and determining their strength	22
Wording of recommendations and of considerations of implementation, feasibility and equity	14
Reporting and peer review	15
Publication of the guideline	13
Dissemination and implementation	13
Evaluation and use	7
Manage alerts for new evidence relating to PICO questions	1
Total	441

For what reason do you use this tool?	Total
To be more efficient	74
To improve collaboration	52
For better version control/historical record	37
For improving the quality of the guidelines	57
For standardisation purposes	45
Total	265

THE UNIVERSITY of ADELAIDE

49

Received: 3 August 2023 | Revised: 15 May 2024 | Accepted: 2 June 2024
DOI: 10.1002/jrsm.1731

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Automation tools to support undertaking scoping reviews

Hanan Khalil[1,2] | Danielle Pollock[3,4] | Patricia McInerney[5] | Catrin Evans[6] | Erica B. Moraes[7,8] | Christina M. Godfrey[9] | Lyndsay Alexander[10,11] | Andrea Tricco[9,12,13] | Micah D. J. Peters[4,14,15] | Dawid Pieper[16,17] | Ashrita Saran[18] | Daniel Ameen[19] | Petek Eylul Taneri[20] | Zachary Munn[3,4]

[1]School of Psychology and Public Health, Department of Public Health, La Trobe University, Melbourne, Australia
[2]The Queensland Centre of Evidence Based Nursing and Midwifery: A JBI Centre of Excellence, Brisbane, Queensland, Australia
[3]JBI, University of Adelaide, Adelaide, Australia
[4]Health Evidence Synthesis, Recommendations and Impact (HESRI), School of Public Health, University of Adelaide, Adelaide, Australia
[5]The Wits JBI Centre for Evidence-Based Practice: A JBI Centre of Excellence, Faculty of Health Sciences, University of the Witwatersrand, South Africa
[6]The Nottingham Centre for Evidence Based Healthcare: A JBI Centre of Excellence, University of Nottingham, UK
[7]Nursing School, Department of Nursing Fundamentals and Administration, Federal Fluminense University, Rio de Janeiro, Brazil
[8]The Brazilian Center of Evidence-based Healthcare: A JBI Centre of Excellence – JBI, Brazil
[9]Queen's Collaboration for Health Care Quality: A JBI Centre of Excellence, Queen's University School of Nursing, Kingston, Ontario, Canada
[10]The Scottish Centre for Evidence-based, Multi-Professional Practice: A JBI Centre of Excellence, Aberdeen, UK
[11]School of Health Sciences, Robert Gordon University, Aberdeen, UK
[12]Epidemiology Division and Institute for Health, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[13]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[14]University of South Australia, Clinical and Health Sciences, Rosemary Bryant AO Research Centre, Adelaide, South Australia, Australia
[15]University of Adelaide, Faculty of Health and Medical Sciences, Adelaide Nursing School, Adelaide, Australia
[16]Faculty of Health Sciences Brandenburg, Brandenburg Medical School (Theodor Fontane), Institute for Health Services and Health System Research, Rüdersdorf, Germany
[17]Center for Health Services Research, Brandenburg Medical School (Theodor Fontane), Rüdersdorf, Germany
[18]Campbell South Asia, India
[19]Faculty of Medicine, Nursing and Health Sciences, School of Medicine, Monash University, Australia
[20]HRB Trials Methodology Research Network, College of Medicine, Nursing and Health Sciences, University of Galway, Galway, Ireland

Correspondence
Hanan Khalil, School of Psychology and Public Health, Department of Public Health, La Trobe University, 360 Collins Street, Melbourne, VIC 3000, Australia.
Email: h.khalil@latrobe.edu.au

Funding information
Tier 2 Canada Research Chair; NHMRC investigator grant, Grant/Award Number: APP1195676

Abstract

Objective: This paper describes several automation tools and software that can be considered during evidence synthesis projects and provides guidance for their integration in the conduct of scoping reviews.

Study Design and Setting: The guidance presented in this work is adapted from the results of a scoping review and consultations with the JBI Scoping Review Methodology group.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Research Synthesis Methods* published by John Wiley & Sons Ltd.

THE UNIVERSITY of ADELAIDE

50

Levels of automation

Level 1: tools to improve the file management process such as citation databases, reference management software and systematic review management software.

Level 2: tools to enable workflow prioritisation such as prioritisation of relevant abstracts.

Level 3: tools used to perform a task automatically but unreliably and require human supervision or otherwise provide the option to override the tools decisions manually.

Level 4: tools that eliminate the need for human participation altogether.

A question of trust: can we build an evidence base to gain trust in systematic review automation technologies?

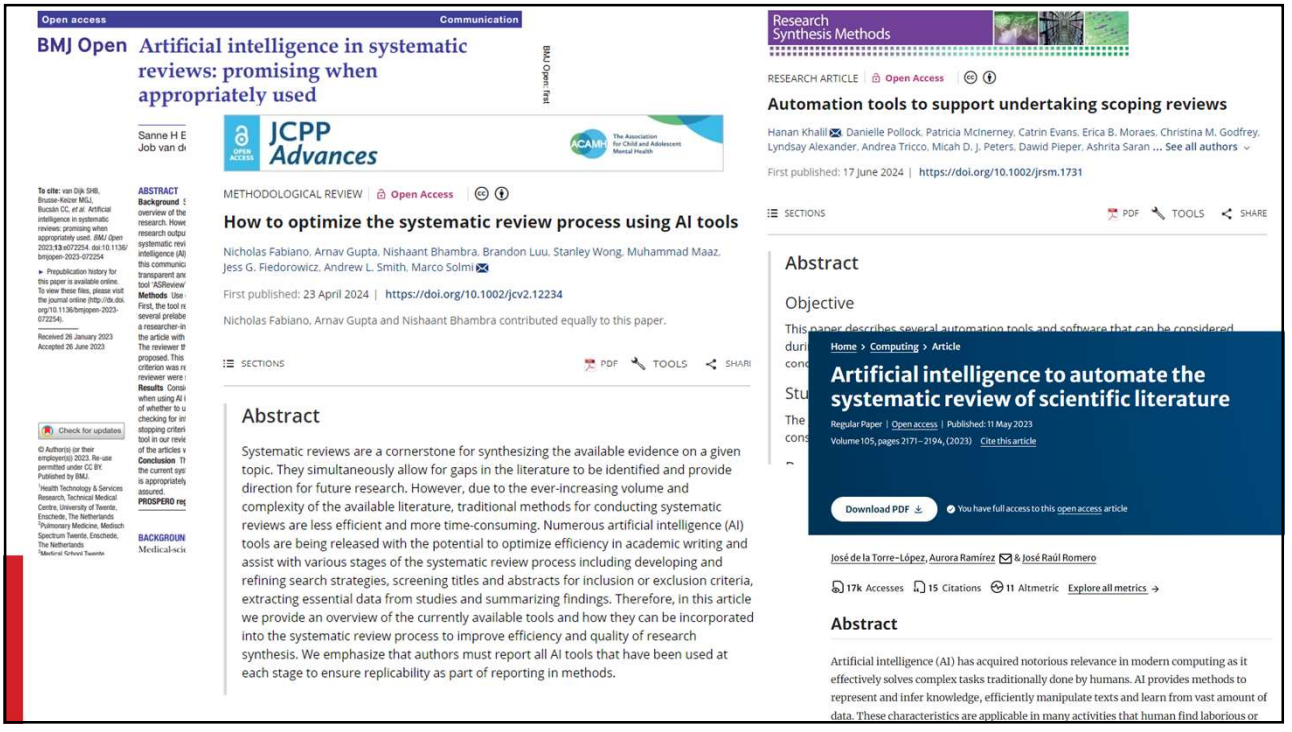
Annette M. O'Connor, Guy Tsafnat, James Thomas, Paul Glasziou, Stephen B. Gilbert & Brian Hutton

Systematic Reviews 8, Article number: 143 (2019) | Cite this article

TABLE 1 Tasks of a scoping review and how automation tools could support their efficiency.

Scoping review task	Method used	Suggested automation tools to increase efficiency or gain	Level of automation	Availability of the tool (open access/proprietary)
Team establishment	Personnel needed for the completion of the review	Crowd sourcing or using task sharing platforms i.e., JBI SUMARI and Covidence. Management teams—Trello	Level 1	Proprietary
Protocol development	Templated reports of some report items	JBI SUMARI and Revman	Level 1	Proprietary
Searches	Running searches on databases	Automatic alerts from databases. Database (i.e., Cochrane, JBI and Campbell, Health Database Advanced Search (HDAS)) Automatic retrieval of full-text papers (e.g., CrossRef), Polyglot Search Translator Litsuggest, eSuRFr	Level 4	CrossRef and Litsuggest (Open access) The rest of the tools are proprietary.
Deduplication	Deleting duplicates	Endnote, Systematic Review Accelerator (SRA De-duplicate), Abstrackr, Rayyan, EPPI-Reviewer, Deduklick, Covidence and DistillerSR.	Level 3	Abstrackr (Open access) Rayyan (Open Access to early career researchers) The remainder of tools are proprietary.
Screening	Selecting studies for inclusion	Crowdsourced inclusion decision and machine learning classifier (Abstrackr, Rayyan, BIBOT, EPPI-Reviewer, ASReview and DistillerSR). Covidence is also used for this step.	Level 2	Abstrackr (Open access) Rayyan (Open Access to early career researchers) The remainder of tools are proprietary.
Data extraction	Extracting information on participants, Concept and context	Machine learning-information extraction systems such as DistillerSR and ExaCT), Tabulizer, UDPipe, Webplotdigitizer. For non-machine learning but wanting a faster extraction process- Covidence	Level 4	UDPipe (open access) and the remainder of the tools are proprietary.
Data charting	Presentation of data	GoogleCharts, Tableau Public, Google Data studio, Drawio, Eppi- reviewer, NVivo.	Level 1	Googlecharts, Google data studio and Drawio (open access). The remainder of the tools are proprietary.
Report writing	Templated reports of some report items	JBI SUMARI, Revman and Robotreviewer	Level 1	Proprietary

THE UNIVERSITY of ADELAIDE



Open access Communication

BMJ Open

Artificial intelligence in systematic reviews: promising when appropriately used

BMJ Open: first

Sanne H E
Job van d

JCPP Advances

METHODOLOGICAL REVIEW | Open Access

How to optimize the systematic review process using AI tools

Nicholas Fabiano, Arnav Gupta, Nishaant Bhambra, Brandon Luu, Stanley Wong, Muhammad Maaz, Jess G. Fiedorowicz, Andrew L. Smith, Marco Solmi

First published: 23 April 2024 | https://doi.org/10.1002/jcv2.12234

Nicholas Fabiano, Arnav Gupta and Nishaant Bhambra contributed equally to this paper.

SECTIONS PDF TOOLS SHARE

Abstract

Systematic reviews are a cornerstone for synthesizing the available evidence on a given topic. They simultaneously allow for gaps in the literature to be identified and provide direction for future research. However, due to the ever-increasing volume and complexity of the available literature, traditional methods for conducting systematic reviews are less efficient and more time-consuming. Numerous artificial intelligence (AI) tools are being released with the potential to optimize efficiency in academic writing and assist with various stages of the systematic review process including developing and refining search strategies, screening titles and abstracts for inclusion or exclusion criteria, extracting essential data from studies and summarizing findings. Therefore, in this article we provide an overview of the currently available tools and how they can be incorporated into the systematic review process to improve efficiency and quality of research synthesis. We emphasize that authors must report all AI tools that have been used at each stage to ensure replicability as part of reporting in methods.

Research Synthesis Methods

RESEARCH ARTICLE | Open Access

Automation tools to support undertaking scoping reviews

Hanan Khalil, Danielle Pollock, Patricia McInerney, Catrin Evans, Erica B. Moraes, Christina M. Godfrey, Lyndsay Alexander, Andrea Tricco, Micah D. J. Peters, Dawid Pieper, Ashrita Saran ... See all authors

First published: 17 June 2024 | https://doi.org/10.1002/jrsm.1731

SECTIONS PDF TOOLS SHARE

Abstract

Objective

Home › Computing › Article

Artificial intelligence to automate the systematic review of scientific literature

Regular Paper | Open access | Published: 11 May 2023
Volume 105, pages 2171–2194, (2023) Cite this article

Download PDF You have full access to this open access article

José de la Torre-López, Aurora Ramírez & José Raúl Romero

17k Accesses 15 Citations 11 Altmetric Explore all metrics

Abstract

Artificial intelligence (AI) has acquired notorious relevance in modern computing as it effectively solves complex tasks traditionally done by humans. AI provides methods to represent and infer knowledge, efficiently manipulate texts and learn from vast amount of data. These characteristics are applicable in many activities that human find laborious or

53

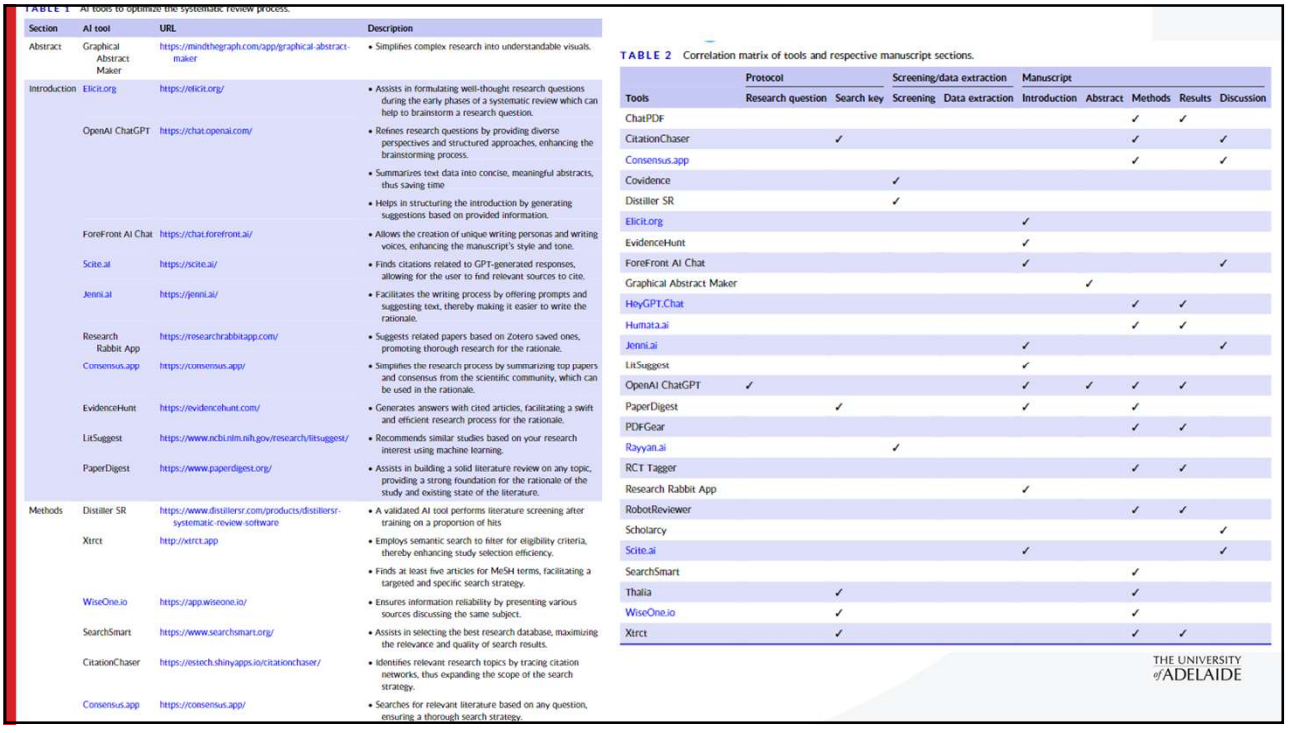


TABLE 1 AI tools to optimize the systematic review process.

Section	AI tool	URL	Description
Abstract	Graphical Abstract Maker	https://mindthegraph.com/app/graphical-abstract-maker	• Simplifies complex research into understandable visuals.
Introduction	Elicit.org	https://elicit.org/	• Assists in formulating well-thought research questions during the early phases of a systematic review which can help to brainstorm a research question.
	OpenAI ChatGPT	https://chat.openai.com/	• Refines research questions by providing diverse perspectives and structured approaches, enhancing the brainstorming process.
			• Summarizes text data into concise, meaningful abstracts, thus saving time
			• Helps in structuring the introduction by generating suggestions based on provided information.
	ForeFront AI Chat	https://chat.forefront.ai/	• Allows the creation of unique writing personas and writing voices, enhancing the manuscript's style and tone.
	Scite.ai	https://scite.ai/	• Finds citations related to GPT-generated responses, allowing for the user to find relevant sources to cite.
	Jenni.ai	https://jenni.ai/	• Facilitates the writing process by offering prompts and suggesting text, thereby making it easier to write the rationale.
	Research Rabbit App	https://researchrabbitapp.com/	• Suggests related papers based on Zotero saved ones, promoting thorough research for the rationale.
	Consensus.app	https://consensus.app/	• Simplifies the research process by summarizing top papers and consensus from the scientific community, which can be used in the rationale.
	EvidenceHunt	https://evidencehunt.com/	• Generates answers with cited articles, facilitating a swift and efficient research process for the rationale.
	LitSuggest	https://www.ncbi.nlm.nih.gov/research/litsuggest/	• Recommends similar studies based on your research interest using machine learning.
	PaperDigest	https://www.paperdigest.org/	• Assists in building a solid literature review on any topic, providing a strong foundation for the rationale of the study and existing state of the literature.
Methods	Distiller SR	https://www.distillersr.com/products/distillersr-systematic-review-software	• A validated AI tool performs literature screening after training on a proportion of hits
	Xtrct	http://xtrct.app	• Employs semantic search to filter for eligibility criteria, thereby enhancing study selection efficiency.
			• Finds at least five articles for MeSH terms, facilitating a targeted and specific search strategy.
	WiseOne.io	https://app.wiseone.io/	• Ensures information reliability by presenting various sources discussing the same subject.
	SearchSmart	https://www.searchsmart.org/	• Assists in selecting the best research database, maximizing the relevance and quality of search results.
	CitationChaser	https://estech.shinyapps.io/citationchaser/	• Identifies relevant research topics by tracing citation networks, thus expanding the scope of the search strategy.
	Consensus.app	https://consensus.app/	• Searches for relevant literature based on any question, ensuring a thorough search strategy.

TABLE 2 Correlation matrix of tools and respective manuscript sections.

	Protocol		Screening/data extraction		Manuscript				
Tools	Research question	Search key	Screening	Data extraction	Introduction	Abstract	Methods	Results	Discussion
ChatPDF							✓	✓	
CitationChaser		✓					✓		✓
Consensus.app							✓		✓
Covidence			✓						
Distiller SR			✓						
Elicit.org					✓				
EvidenceHunt					✓				
ForeFront AI Chat					✓				✓
Graphical Abstract Maker						✓			
HeyGPT.Chat							✓	✓	
Humata.ai							✓	✓	
Jenni.ai					✓				✓
LitSuggest					✓				
OpenAI ChatGPT	✓				✓	✓	✓	✓	
PaperDigest		✓			✓		✓		
PDFGear							✓	✓	
Rayyan.ai			✓						
RCT Tagger							✓	✓	
Research Rabbit App					✓				
RobotReviewer							✓	✓	
Scholarcy									✓
Scite.ai					✓				✓
SearchSmart							✓		
Thalia		✓					✓		
WiseOne.io		✓					✓		
Xtrct		✓					✓	✓	

THE UNIVERSITY of ADELAIDE

54

ecdc

Contents

Abbreviations iv
Glossary iv
Executive summary 5
Introduction 5
Study aims and methods 5
Results 5
Conclusions 7
Study objectives 8
Research methods 9
Task 1: Systematic review 9
Task 1.2 Screening 9
Task 1.3 Data extraction 9
Task 1.4 Analysis and synthesis 10
Task 2: Qualitative data collection (mapping exercise) 10
..... 11
..... 12
..... 12
..... 14
..... 15
..... 15
..... 16
..... 17
..... 17
..... 17
..... 18
..... 18
..... 18
..... 19
..... 20
..... 20
..... 20
..... 21
..... 22
..... 22
..... 22
..... 23
..... 23
..... 24
Current approaches to evidence syntheses and changes during the pandemic 24
Challenges of conventional literature review 24
Extent of the use of automation 25
Software suites 25
Advantages of using automated approaches to conduct evidence syntheses 26
Reduction in burden 26
Ease of use 26
Enables completion of new tasks 27
Challenges and needs in using automation for evidence syntheses 27
Continued need for human input 27
Technical limitations 27
Need for staff training and capacity 28
Cost and equipment/resource requirements 28
Lack of trust and performance evaluation 28
Factors that can support the use of automation 29
Collaboration 30
Discussion and conclusions 31
References 34
Annex 1 Participants in focus groups, interviews and survey 40

Results

Based on the qualitative data collected from EU/EEA public health bodies participating in this study, we found that automated technology is used to a limited extent for conducting evidence synthesis, with most participants not currently using these types of tools. This may be because they do not conduct any type of review within their organisation (instead, relying on reviews produced by others) or because of different challenges and obstacles that can arise when attempting to use automated approaches (discussed below). However, there were a few public heath bodies that used automated tools, sometimes extensively. Among these public health bodies, technologies were used in different stages of the evidence synthesis process, although they seem to be particularly used to semi-automate the screening stage of reviews. For other public health bodies, the adoption of new technologies was being explored for the future.

Use and impact of new technologies for evidence synthesis

Literature review and qualitative data collection

2022 (Disclosue: I was a survey participant)

THE UNIVERSITY of ADELAIDE

55

AHRQ Agency for Healthcare Research and Quality

Search All AHRQ Site

Effective Health Care Program

Powered by the Evidence-based Practice Centers

Search Evidence Reports

Evidence Reports | Health Topics | Provide Input | Research Methods & Tools | About EHC | Contact EHC

Home > Evidence Reports

Key Question (KQs)

Tools t

KQ 1: What tools that use artificial intelligence (AI) or machine learning (ML) to semi- or fully automate any aspect of evidence synthesis product (ESP) production have had their performance evaluated?

KQ 1a: How do these tools perform?

Research Protocol | Apr 15, 2024

Download Full Content

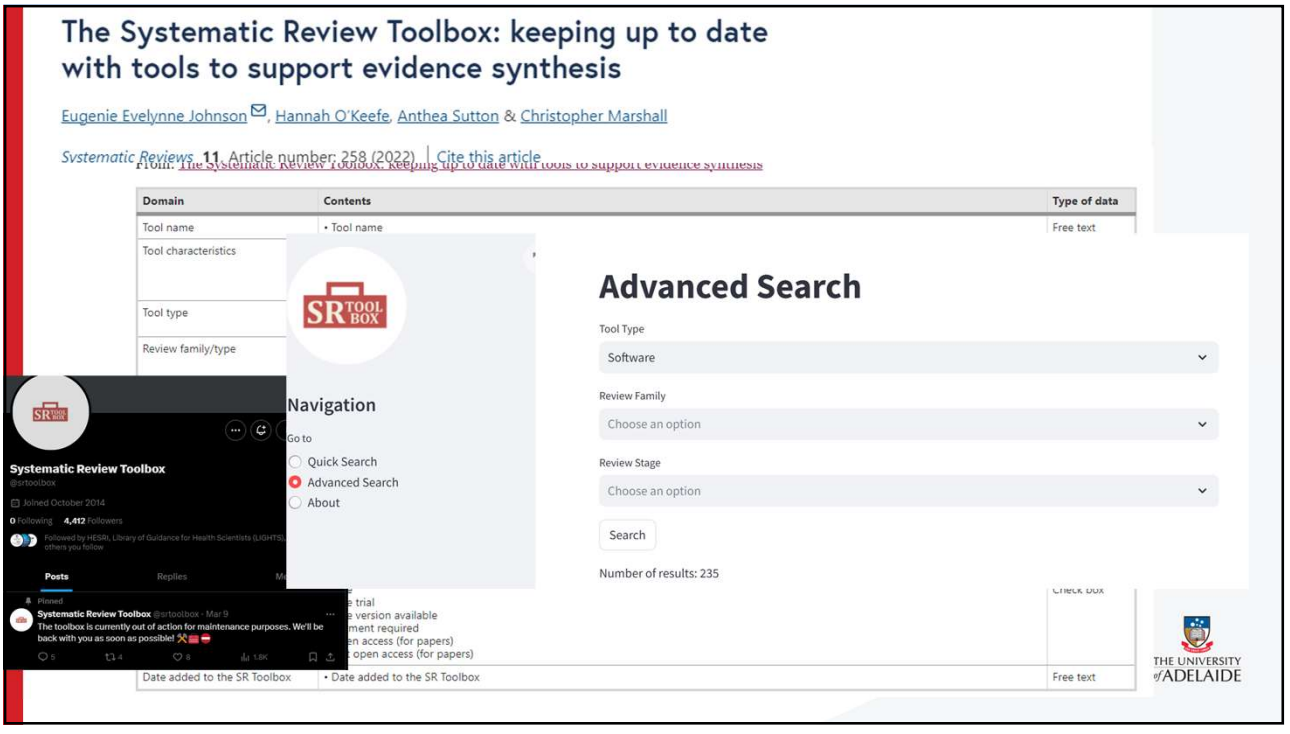
Page Contents

- I. Background and Purpose of the Review
- II. Key Questions and Eligibility Criteria
- III. Methods
- IV. References
- V. Definition of Terms and Abbreviations
- VI. Summary of Protocol Amendments
- VII. EPC Team Disclosures
- VIII. Role of the Funder
- IX. Registration

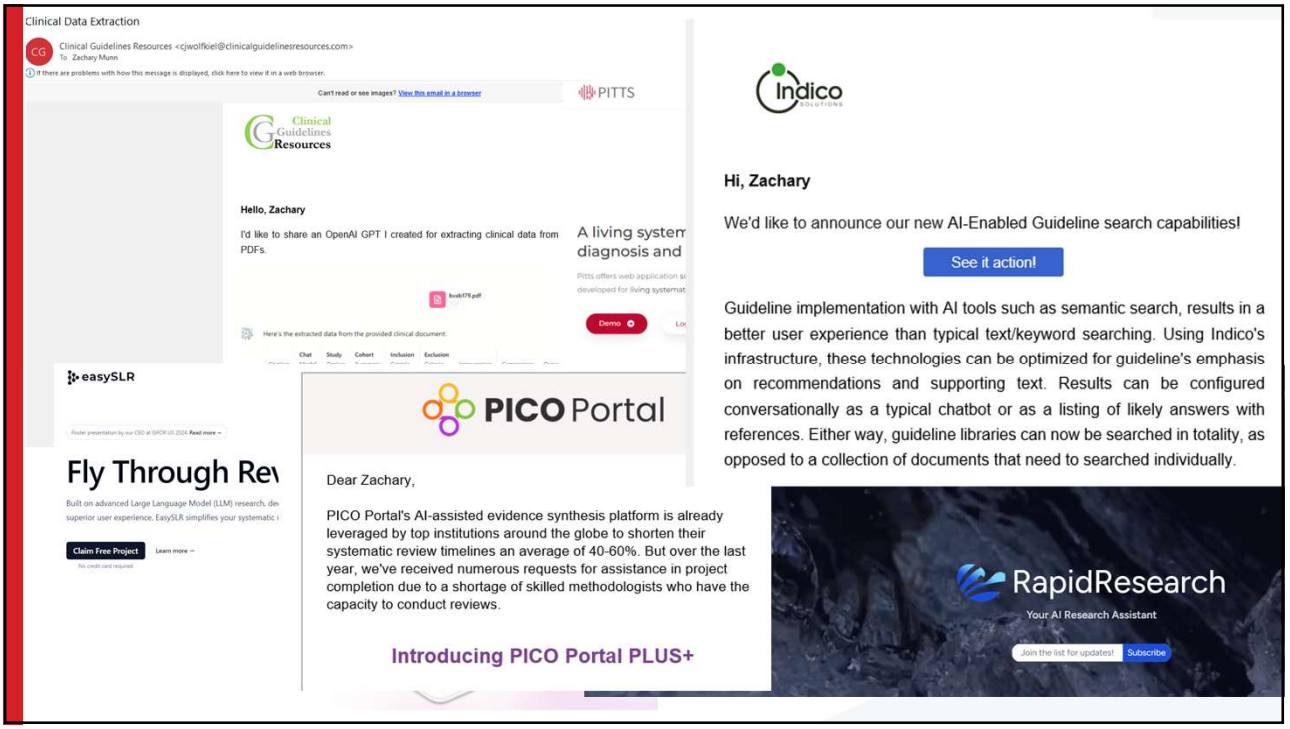
MACHINE LEARNING

THE UNIVERSITY of ADELAIDE

56


The Systematic Review Toolbox: keeping up to date with tools to support evidence synthesis
Eugenie Evelynne Johnson, Hannah O'Keefe, Anthea Sutton & Christopher Marshall
Systematic Reviews 11, Article number: 258 (2022) | Cite this article
From: The Systematic Review Toolbox: keeping up to date with tools to support evidence synthesis
Domain
Contents
Type of data
Tool name
• Tool name
Free text
Tool characteristics
Tool type
Review family/type
Date added to the SR Toolbox
• Date added to the SR Toolbox
Free text
Navigation
Go to
Quick Search
Advanced Search
About
Advanced Search
Tool Type
Software
Review Family
Choose an option
Review Stage
Choose an option
Search
Number of results: 235
Systematic Review Toolbox
@srtoolbox
Joined October 2014
0 Following 4,412 Followers
Followed by HESRI, Library of Guidance for Health Scientists (LIGHTS), others you follow
Posts
Replies
Pinned
Systematic Review Toolbox @srtoolbox · Mar 9
The toolbox is currently out of action for maintenance purposes. We'll be back with you as soon as possible!
THE UNIVERSITY of ADELAIDE

57


Clinical Data Extraction
Clinical Guidelines Resources <cjwolfkiel@clinicalguidelinesresources.com>
To Zachary Munn
Can't read or see images? View this email in a browser
PITTS
Clinical Guidelines Resources
Hello, Zachary
I'd like to share an OpenAI GPT I created for extracting clinical data from PDFs.
A living system diagnosis and
Demo
Here's the extracted data from the provided clinical document:
Chat
Study
Cohort
Inclusion
Exclusion
easySLR
Fly Through Rev
Claim Free Project
Learn more
PICO Portal
Dear Zachary,
PICO Portal's AI-assisted evidence synthesis platform is already leveraged by top institutions around the globe to shorten their systematic review timelines an average of 40-60%. But over the last year, we've received numerous requests for assistance in project completion due to a shortage of skilled methodologists who have the capacity to conduct reviews.
Introducing PICO Portal PLUS+
Indico
Hi, Zachary
We'd like to announce our new AI-Enabled Guideline search capabilities!
See it action!
Guideline implementation with AI tools such as semantic search, results in a better user experience than typical text/keyword searching. Using Indico's infrastructure, these technologies can be optimized for guideline's emphasis on recommendations and supporting text. Results can be configured conversationally as a typical chatbot or as a listing of likely answers with references. Either way, guideline libraries can now be searched in totality, as opposed to a collection of documents that need to searched individually.
RapidResearch
Your AI Research Assistant
Join the list for updates!
Subscribe

58

From:
Date: 17 June 2026 at 10:16:17 am AEST
To: Zachary Munn
Subject: evidence synthesis perspective

CAUTION: External email. Only click on links or open attachments from trusted senders.

Hey Zachary,

You run HESRI at Adelaide Uni, so you've probably seen the full range of systematic reviews across health. We built an AI platform for SRs and validated it against Cochrane (published on medRxiv), but I'm sure there are gaps only someone with your experience would catch.

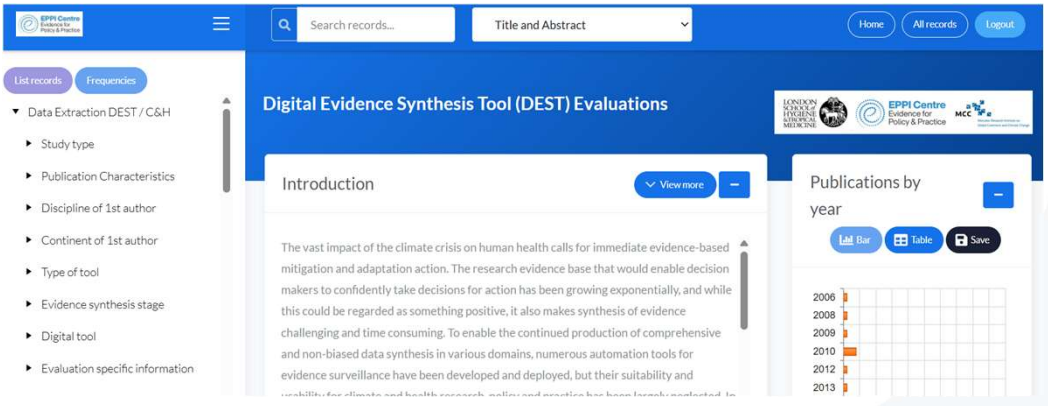
Your role as director of evidence synthesis means you think about the whole process, not just one step. I'd love to hear your honest take on where AI could actually help (and where it can't). Would you be up for a quick chat sometime?

THE UNIVERSITY of ADELAIDE

59

Digital Evidence Synthesis Tools (DEST)

https://eppi.ioe.ac.uk/eppi-vis/Frequencies/GetMapById



THE UNIVERSITY of ADELAIDE

60

Critical questions to ask when considering using a new tool for evidence synthesis

Does it enable me to draw on the sum of current knowledge?

Or does it present an incomplete or biased picture?

Does it enable me to distinguish between reliable and unreliable research?

Or does it treat all research as equally reliable?

Is the tool itself reliable?

Do we know how it works? Is it validated for my use case?

Slide from a presentation by James Thomas:
https://eppi.ioe.ac.uk/CMS/Portals/0/ARC%20masterclass%20on%20AI%20for%20web2.pdf

61

Automation tools to support undertaking scoping reviews

The International Collaboration for Automation of Systematic Reviews (ICASR) has put forward a set of principles for researchers to use automation for systematic reviews highlighting that automation has the potential to facilitate the production of systematic reviews that adhere to high standards for the reporting, conduct, and updating of rigorous reviews.

The Vienna Principles

1. Systematic reviews involve multiple tasks, each with different issues, but all must be improved.
2. Automation may assist with all tasks, from scoping reviews to identifying research gaps as well protocol development to writing and dissemination of the review.
3. The processes for each task can and should be continuously improved, to be more efficient and more accurate.
4. Automation can and should facilitate the production of systematic reviews that adhere to high standards for the reporting, conduct and updating of rigorous reviews.
5. Developments should also provide for flexibility in combining and using, e.g. subdividing or merging steps and allowances for different users to use different interfaces.
6. Different groups with different expertise are working on different parts of the problem; to improve reviews as a whole will require collaboration between these groups.
7. Every automation technique should be shared, preferably by making code, evaluation data and corpora available for free.
8. All automation techniques and tools should be evaluated using a recommended and replicable method with results and data reported.

ICASR

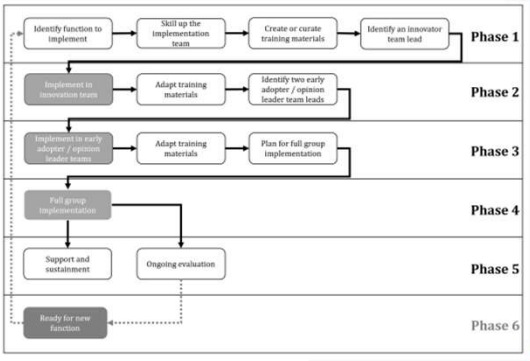
International Collaboration for the Automation of Systematic Reviews

THE UNIVERSITY of ADELAIDE

62

This report provides recommendations on how to carry out implementation of ML functions in ML-naïve evidence synthesis groups, based on the experiences implementing ML at HTV. It offers "best practice" suggestions, rooted in our reflections on implementation, aiming to assist other ML naïve groups or institutions in implementing ML functions in the evidence synthesis process. The guide is adaptable to different organizational goals and objectives, while providing insights applicable to implementation of various ML tools and functions.

The report is structured into three main sections corresponding to different phases of implementation: pre-implementation, implementation, and sustainment/evaluation. We use the EPIS framework throughout the document as a tool to explain the different implementation phases and important aspects to consider in each phase. Each section concludes with a "Take home message" based on our implementation experiences, summarized as practical tips on important aspects that we believe are important to consider in the implementation process.



THE UNIVERSITY of ADELAIDE

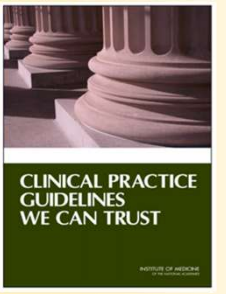
63

Developing guidelines is a human endeavour

THE UNIVERSITY of ADELAIDE

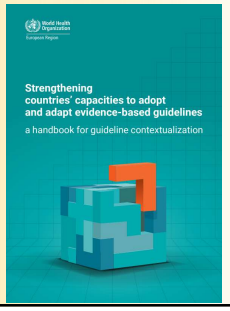
64

Guidelines



> Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a **systematic review** of evidence and an assessment of the **benefits and harms** of alternative care options.

(Institute of Medicine, 2011)



> Trustworthy guidelines are the cornerstone of giving evidence-informed advice in response to clinical, public health and health policy questions.
> They address those questions by using the **best available evidence** and transparently integrating the **judgements** of **experts** and the **input of stakeholders** in the process

(WHO Handbook for guideline contextualization, 2023)

65

How AI works

Generative AI (LLMs) provide a statistical representation of language – in contrast to human reasoning

Rule based calculations

Don't always like uncertainty, nuance

Unclear decisions

Makes up fake information/hallucinates

Perpetuates structural bias, stereotypes, etc

In guidelines:

Evidence AND value based judgements

Conscious dealing with uncertainty, conflicting results, trade offs

Referencing, cited and transparent decisions

Acknowledge and integrate different perspectives

Recognise importance of human connection, patients as persons, lived experience

THE UNIVERSITY of ADELAIDE

66

Slide courtesty of Valentin III Dones from GES*

*Note appropriate recognition of borrowed ideas

THE UNIVERSITY of ADELAIDE

67

What if we do it once and do it right

Computable data

RoB

Results data coded

All EtD criteria coded and available

SRDR, Cochrane database, FEVIR, ALIVE, ESIC

Evidence inventories/registries, recommendation maps, adolopment

THE UNIVERSITY of ADELAIDE

68

The future…..

Never a better time for automation in evidence synthesis due to:

- Open science and reusable data
- Modern technology
- Advances in AI
- Tech literacy
- Sharing and collaboration
- Interoperability and common standards
- Societal needs and pressures requiring innovation

Never a worse time for automation in evidence synthesis due to:

- Contributing to more poor quality reviews
- Competition
- Risks of AI not being addressed
- Tools appearing and disappearing
- Societal pressures leading to shortcuts
- Lack of commitment to common standards

THE UNIVERSITY of ADELAIDE

69

Where does this leave us?

Lots of hype but not sure all applications 'stack up'

Everyone is looking for an easy answer – I don't think it exists (yet)

Personally – more trust in credible, evidence synthesis specialist organisations/collaboratons – i.e. Covidence, Eppi-Reviewer, SR Accelerator, MAGICapp, GRADEPro

Less likely to disappear and more likely to be maintained – or provide support

These will make improvements and save time – but may not be 'game changing'

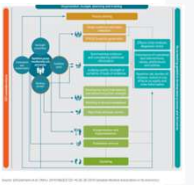
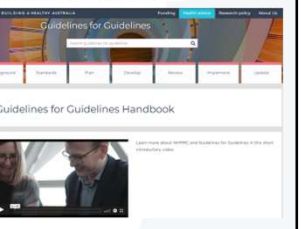
….we need a living evidence review on SR automation tools (plus evaluations) to underpin a resource such as the SR toolbox or ICASR repository (and collaboration in tool development)

THE UNIVERSITY of ADELAIDE

70

Ensuring high quality and efficient development of trustworthy guidelines

- Planning and production - GIN McMaster checklist
- Handbooks and guidance – NHMRC, GRADE
- Reporting standards – RIGHT
- Evaluate trustworthiness of a guideline – AGREE
- Software – MAGICApp, GRADEPro, Covidence
- Networking, registering guidelines, collaboration - GIN
- Adapt and adopt – GRADE Adolopment
- Engagement with community – MuSE/ GIN Public
- For education and training – INGUIDE

GRADE

WHO Handbook for Guideline Development

RIGHT
Reporting Items for Practice Guidelines in Healthcare

AGREE II

M Authoring & Publication Platform

covidence

GIN Guidelines International Network

GIN Public

MuSE

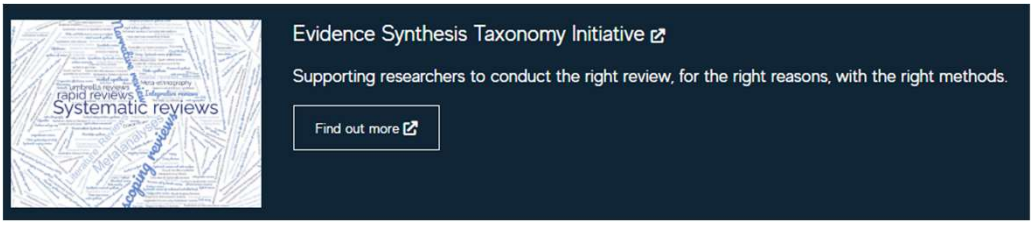
INGUIDE International Guideline Development Credentialing & Certification Program

THE UNIVERSITY of ADELAIDE

Concluding thoughts

- No single AI/ML software solution – yet(?)
- Consider evidence synthesis type, scope, restrictions
- AI (and computable evidence) will likely let us do synthesis ‘at the push of the button’ or a form of it at some stage
- Guidelines still require the human touch
- Need to consider people, processes, technology
- Need to invest in training and upskilling
- Build efficient workflows for different projects through testing and refinement
- Consider other approaches to improve efficiency (living, mapping, adolopment, updating, centralised registry, collaboration [built on FAIR/open science/standards for interoperability])
- We have come a long way (paper, manual referencing, excel, no interoperability)– but still have further to go!

THE UNIVERSITY of ADELAIDE



Right Reviews
Right Reasons
Right Methods

THE UNIVERSITY of ADELAIDE

73

ChatGPT/ Claude AI

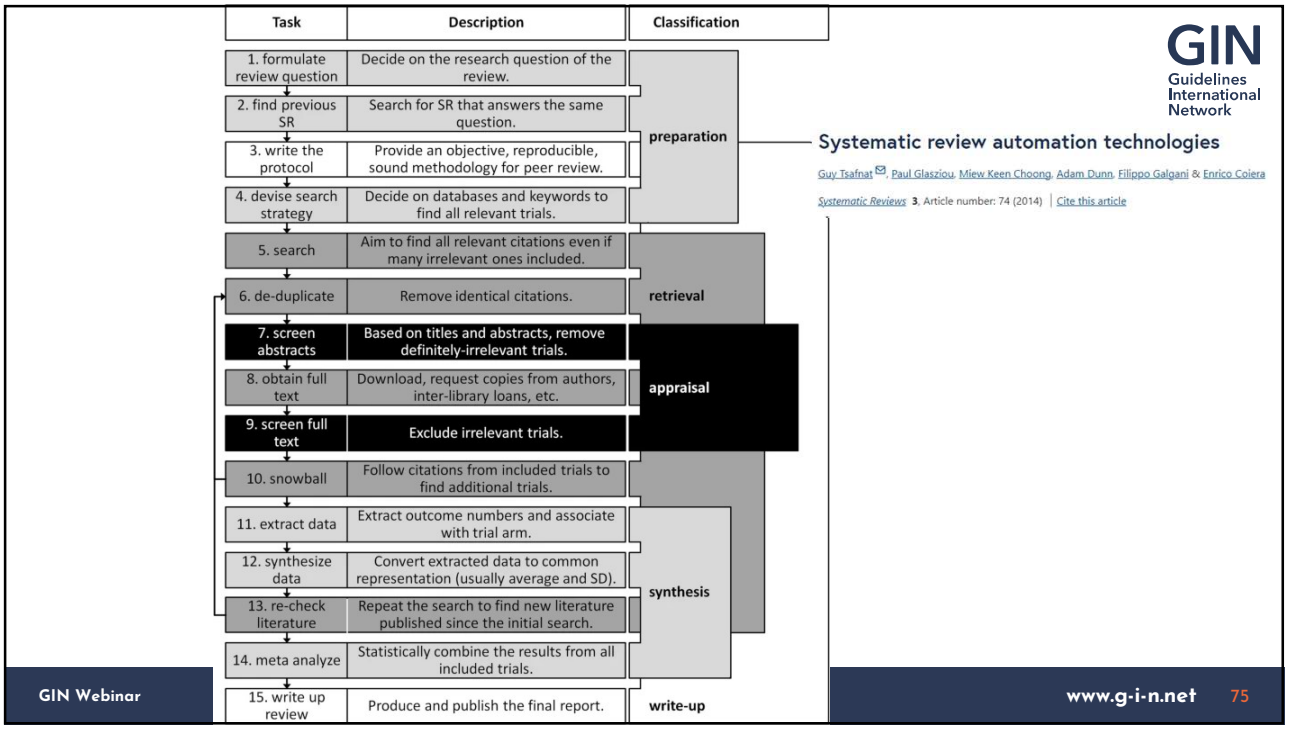
When does it work? How do we use it?

- **Background and discussion ideas**
- **Risk management/project management plans**
- **Consider coverage of all knowledge, credibility of knowledge**
- **Suggested text for some methods sections**
- **Ideas on risk of bias/ appraisal but not trustworthy *(Claude impressive)***
- **Search term generation/ search strings (but not trustworthy)**
- **R code (?)**
- **Plain language summary/ executive summary**

- **What will it be able to do in the future? Will this be the 'game changer'?**

THE UNIVERSITY of ADELAIDE

74



75

Workflows: SR of RCTs

Protocol: Templates – SUMARI, RevMan or methods wizard

Searching: Polyglot search translator, Citation chaser/SpiderCite, Co-cites, De-duplicator, Endnote

Screening: Covidence (including RCT classifier)

Extraction: Covidence

RoB: Covidence, RoB excel

Analysis: RevMan

Write up: Word, endnote, Revman

Certainty of findings: GRADEPro

THE UNIVERSITY of ADELAIDE

76

Workflows: Scoping Review/Evidence Maps

Protocol: Templates, SUMARI

Searching: Polyglot search translator, Citation chaser/SpiderCite, Co-cites, De-duplicator

Screening: Covidence/Eppi-Reviewer

Extraction: Covidence/Eppi-Reviewer

RoB: NA

Analysis/Presentation: Covidence/Eppi-Reviewer/Eppi-Mapper/Eppi-Visualiser/Nvivo

Write up: Word, endnote

Certainty of findings: NA

THE UNIVERSITY of ADELAIDE

Workflows: Guidelines

As above, but also:

Facilitate report writing, publishing, living mode: MAGICapp

Panel voting, judgements, reviewing: GRADEPro, MAGICapp

Automatically generated shared decision making tools: GRADEPro, MAGICapp

THE UNIVERSITY of ADELAIDE