

The benefits of data sharing and ensuring open sources of systematic review data

Sze Lin Yoong^{1,2,3,4,5}, Heidi Turon^{3,4,5}, Alice Grady^{2,3,4,5}, Rebecca Hodder^{2,3,4,5}, Luke Wolfenden^{2,3,4,5}

¹Faculty of Health, Arts and Design, Swinburne University of Technology, John Street, Hawthorn, VIC 3122, Australia

²Hunter New England Population Health, Longworth Avenue Wallsend, NSW 2287, Australia

³School of Medicine and Public Health, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

⁴Priority Research Centre in Health Behaviour, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

⁵Hunter Medical Research Institute, Kookaburra Circuit, New Lambton Heights, NSW 2305, Australia

Address correspondence to Sze Lin (Serene) Yoong, Email: syoong@swin.edu.au.

ABSTRACT

Aims The benefits of increasing public access to data from clinical trials are widely accepted. Such benefits extend to the sharing of data from high-quality systematic reviews, given the time and cost involved with undertaking reviews. We describe the application of open sources of review data, outline potential challenges and highlight efforts made to address these challenges, with the intent of encouraging publishers, funders and authors to consider sharing review data more broadly.

Results We describe the application of systematic review data in: (i) advancing understanding of clinical trials and systematic review methods, (ii) repurposing of data to answer public health policy and practice relevant questions, (iii) identification of research gaps and (iv) accelerating the conduct of rapid reviews to inform decision making. While access, logistical, motivational and legal challenges exist, there has been progress made by systematic review, academic and funding agencies to incentivise data sharing and create infrastructure to support greater access to systematic review data.

Conclusion There is opportunity to maximize the benefits of research investment in undertaking systematic reviews by ensuring open sources of systematic review data. Efforts to create such systems should draw on learnings and principles outlined for sharing clinical trial data.

Keywords data sharing, open access, systematic reviews

It has been argued that scientists have an ethical and scientific imperative to share data collected through their research.^{1,2} The benefits of increasing public access to data from research activity are well described. Data sharing maximizes transparency and public accountability, increasing the quality of research and reducing researcher burden.³ Sharing data from original research encourages repurposing and exploration of new lines of inquiry, can support greater research output and reduces research waste. As such, calls for data sharing have come from a range of international organizations (including the World Health Organization,⁴ Organization for Economic Co-operation and Development)⁵ and research funders (Bill and Melinda Gates Foundation,⁶ United Kingdom Medical Research Council).⁷ The sharing of clinical trial data is also increasingly a requirement for publication in many high-impact medical journals such as *The BMJ* and *The Lancet*.⁸

While calls for greater data sharing have focused on clinical trials, there may be similar advantages to sharing data collected as part of systematic reviews.⁹ Systematic reviews “seek to collate evidence that fits pre-specified eligibility criteria in order to answer a specific research question. They aim to minimize bias by using explicit, systematic methods documented in advance with a protocol”.¹⁰ A variety of screening and data extraction tasks are undertaken as part of the review process, taking significant time and resulting in a large amount of secondary data being extracted. Each review process generates individual pieces of data—ranging from search strategies,

Sze Lin Yoong, Associate Professor

Heidi Turon, Postdoctoral Researcher

Alice Grady, Research Fellow

Rebecca Hodder, Research Fellow

Luke Wolfenden, Professor

lists of eligible studies (together with their linked publications), details on study characteristics (participants, intervention, details of controls if relevant and outcomes), data used for analysis (effect sizes, variability estimates and meta-analysis outcomes) and risk of bias assessments and summaries of evidence (such as GRADE: Grading of Recommendations, Assessment, Development and Evaluations that provides an overarching grading of evidence to inform clinical practice). Each of these processes, requiring considerable time and resources to complete, produces data that could be applied to generate new knowledge (Table 1). It is likely that some data within reviews may be more broadly available than others, particularly for those outlined within the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance for reporting of systematic reviews.¹¹ For example, a scoping review examining adherence to the PRISMA guidelines found that just over 50% of reviews reported their search strategy in sufficient detail and almost two-thirds reported risk of bias data for individual studies.¹² The extent of availability of data related to analysis of individual's studies (e.g. effect sizes, measures used to calculate standardized effects), however, is limited, with one study reporting that 21% of reviews funded by the Agency for Healthcare Research and Quality (AHRQ) were shared in an open repository, despite being a funding requirement of the agency.¹³

This manuscript aims to describe broad application of open sources of systematic review data, outlines potential barriers to sharing of data and highlights efforts that have been made to address these challenges to date. The intent is that by drawing attention to such issues, this may encourage publishers, funders and authors of reviews to consider implementing strategies to support sharing of review data.

Firstly, we start by discussing how data extracted from systematic reviews can be repurposed to address research and public health policy needs.

a) *To advance understanding of clinical trial and systematic review methods*

Data from systematic reviews have been used to develop research and review methods. For example, secondary data analysis was undertaken to assess the impact of specific biases that form the Cochrane risk of bias tools.^{15,16} Tierney and Stewart¹⁵ used 14 meta-analyses of individual patient data from 133 randomized controlled trials (RCTs) and 21 905 patients to explore how and whether the exclusion of participants post-randomization may impact on outcomes to inform the development of the Cochrane risk of bias tool. When comparing meta-analyses of results the authors found that studies that had excluded patients from analyses were more likely to report findings that favoured the treatment.¹⁵

Additionally, studies have re-analysed meta-analysis data to identify tools that can help explain study variations in review findings, providing insights into trial design characteristics that can influence effect sizes of pooled analyses. For example, a tool that classifies RCTs as pragmatic or explanatory (PRECIS-2) was included as a covariate in meta-regression models and found larger pooled estimate effect sizes for a systematic review of childhood obesity prevention interventions.¹⁶

b) *Repurposing of data to answer public health policy and practice relevant questions*

Data synthesized within systematic reviews can be repurposed and used to answer secondary research questions. For example, a review examining the effectiveness of smoking interventions delivered by health care providers included secondary analysis on trial data from a subset of studies included in a Cochrane review to describe the potential effectiveness of interventions delivered by usual health service staff in reducing child exposure to tobacco smoke.¹⁷ This was used to examine the differential effectiveness of intervention delivered by routine health professionals in reducing tobacco smoke exposure in young children. The secondary analysis and update found no significant effect of health care professional delivered interventions on the primary outcome of reducing tobacco smoke exposure, in contrast to another review¹⁸ that found evidence of effectiveness. These findings highlighted the need for health providers to be supported to implement the interventions with sufficient fidelity and were used to inform development of support strategies and service delivery decisions.

c) *Identification of new research gaps*

Knowledge synthesis, systematic reviews and scoping reviews are primary methods of identifying research gaps.¹⁹ Making data generated by these reviews freely available can allow for the examination of research gaps not explored by primary authors and enable the conduct of subgroup analyses to identify potential interactions and trends.²⁰ For example, the Cochrane Dementia and Cognitive Improvement Group have called for expressions of interest to repurpose systematic review data from reviews published by this group to answer clinical questions on a related topic of vascular dementia research.²¹ Using reviews in this way can help identify promising research avenues more rapidly to justify further efforts for primary research.

d) *Accelerate the conduct of rapid reviews to inform public health decision making*

Rapid reviews are defined as 'a form of knowledge synthesis in which components of the systematic review process

Table 1 Review stage, availability and potential utility of the data gathered from each step

| <i>Step</i> | <i>Available data</i> | <i>Examples of potential utility of the data</i> |
|-----------------------------------|---|--|
| Execute search strategy | All studies meeting search criteria and their source including grey literature searches and information obtained from authors | Developing and testing search filters for new areas of research and refining machine learning processes. |
| Title and abstract screening | All potentially eligible and ineligible studies | Lists of studies screened on broad eligibility criteria could be rescreened to identify specific studies that may meet a narrower inclusion criterion, reducing time spent |
| Full text screening | Included studies together with linked publications Excluded studies, reasons for exclusion | developing search strategies, screening, obtaining full text and linking related publications. ¹⁴ |
| Data extraction | Study, participant, setting characteristics; intervention details; description of outcome measures; study findings (and possible effect sizes and measures of variance), additional unpublished information obtained from authors of included studies | Provides opportunities for new lines of inquiry including: 1. Advancing understanding of clinical trial and systematic review methods. 2. Repurposing of data to answer secondary policy-relevant questions 3. Identification of research gaps 4. Accelerate the conduct of rapid reviews As most of the data is collected in this phase, specific examples are provided in the manuscript below. |
| Risk of bias screening | Sources of potential bias in included studies, study quality | Provides an indication of biases and study quality to support interpretation of outcomes and identify methodological areas of improvement. |
| Assessment of quality of evidence | Certainty of evidence for outcome/s given synthesis of study findings | |

are simplified or omitted to produce information in a timely manner.¹⁶ Rapid reviews have considerable potential to influence practice and policy due to their ability to provide timely information to address targeted questions. This has been witnessed most recently through Cochrane's establishment of a rapid review process for questions relevant to clinical and public health policy for COVID-19. Rapid reviews published to date include a review of school-based measures to contain the pandemic²² and the use of universal screen for COVID-19 infection.²³ These reviews have been used by clinicians, governments and public health advisors to guide the pandemic response.²⁴ In a rapid review, intentional design decisions are made to reduce the overall time taken to identify the relevant literature and provide a synthesis to address a particular research question. Existing systematic reviews are often used as a starting point for rapid reviews; however, several studies suggest that conclusions from rapid reviews can differ from full systematic reviews.²⁵ Making data freely available from existing high-quality reviews may increase the quality of rapid reviews to provide both accurate and timely information to policymakers and end-users.²⁶ Data extracted from new studies can then be combined with previously extracted data to generate more precise estimates where uncertainties still

exist and/or allow for examination of differential intervention effects by subgroups.

Infrastructure and initiatives that facilitate sharing

Although there are significant benefits of systematic review data sharing, several challenges exist including: review access (e.g. located behind a paywall), logistical issues (e.g. lack of metadata, standardized reporting of measures, challenges with accessibility and data quality), motivational barriers (e.g. lack of author incentives), legal impediments (e.g. issues around ownership of data, intellectual property [IP]) and ethical concerns (e.g. appropriate application of authorship guidelines).²⁷

Cochrane and other systematic review agencies have introduced initiatives that support review data-sharing. For example, Cochrane is committed to being fully open access by 2025.²⁸ The creation of central platforms that store lists of studies and standardized data extracted such as the Cochrane Central Register of Controlled trials (CENTRAL) from studies may also address some of the logistical barriers to accessing and sharing data. The AHRQ funds the Systematic Review Data Repository (SRDR)—a free, open-source

platform that houses hundreds of reviews together with downloadable data extraction and risk of bias assessments (<https://srdp.ahrq.gov/projects/published>).¹³ Additionally, there are tools that have been developed that are capable of interrogating and extracting data from existing repositories such as PROSPERO or Open Science Framework where systematic review protocols are routinely registered to reduce duplication.

Cochrane has also curated review data capture in a standardized format that can facilitate data use. For reviews published with Cochrane, data extracted from individual studies are stored in a pre-defined, technical, semi-structured format specific to RevMan (the program used to manage Cochrane reviews). Typically, this is limited to data contained within a single review, however, academics and reviewers can request meta-data from Cochrane for multiple reviews without charge. Various tools have also been designed to automate the process of converting Cochrane meta-data into different forms to facilitate rapid and aggregate use. Examples of such software include those that are designed to ‘sweep’ the RevMan library, clean and tidy the data and generate single files that allow for ease of use (e.g. RAPTOR—RevMan Parsing Tool for Reviewers).²⁹ Although such tools are useful to increase access to data, many are still being tested and consistent reporting and extraction of data within reviews continue to be crucial to allow for better use of shared data. The SRDR provides a standardized data collection form to facilitate standardized reporting; however, over a third of the data uploaded on this platform do not use this form.¹³

Lastly, there is also a need to address the lack of incentives for review teams to undertake data sharing and potential ethical/legal issues. Leading international organizations including the Institute of Medicine (IOM) now endorse the sharing of clinical trials data as ‘duty of care’ governed by principles of overall maximizing of health benefits and a ‘right to science’. Given the substantial outlay of resources for the initial systematic review, due acknowledgement of the original review authorship team is both appropriate and may incentivise efforts by original review authors to share data. Many have called for the need to establish adequate systems distinguishing recognition, ownership and attribution, so that the due acknowledgement can be made as needed.³⁰ The inclusion of DataCite DOIs such as that adopted by SRDR is one way of doing so and allows for the citation and online tracking of shared records. The Cochrane data download policies state that authors should ‘respect and acknowledge the source of the data’,³¹ although no separate DOI is available for data within reviews specifically. Others have proposed establishing separate guidance for data authorship, distinct from authorship of the primary publication that takes into account the different types of data generated within system-

atic review distinguishing between those generated as part of the systematic review process (e.g. search strategy, lists of eligible studies, risk of bias) and those extracted from the original sources (e.g. participant details, results). Given laws around data ownership and copyright particularly pertaining to those extracted from other sources, this should include the application of IP laws to data access arrangement and cover benefit sharing and IP issues transparently.

Journal publishers and research funders can play a key role in promoting data sharing of all types, including that generated by systematic reviews, by mandating that data are shared on available repositories and linked to published protocols (where available), and ensuring publications adhere to reporting guidelines. BioMed Central who publish a broad range of medical journals (encompassing journals such as *Systematic Reviews*, *BMC Public Health*, *BMC Medicine*) have clear policies facilitating open data by applying the Creative Commons CC0 waiver, allowing data to be reused without breaching copyright.³² Several funding agencies including the US National Institute of Health³³ and the Council of European Union,³⁴ and journals such as *The BMJ*³⁵ strongly encourage the sharing of clinical trial data. These policies should similarly extend to systematic review data. Such requirements are starting to emerge for review data, for example, the AHRQ mandates that review data be shared via the SRDR, with reviews funded by the agency making up almost 70% of all repository records.

There has been significant progress in data sharing practices for clinical trials since the publication of the IOM consensus study³⁶ and the publication from Tudur Smith *et al.*³⁷ outlining good practice principles for data sharing that could apply to data from systematic reviews. For example, recommendations around the development of a clear and rigorous system to assess data requests and grant access, as well as establishing data sharing agreements have clear transferability, and can aid in addressing some of the legal challenges that may arise with review data sharing and reduce the burden on researchers to share and use this data. Various producers of systematic reviews including Cochrane have been a strong proponent of open science and data sharing, particularly in the context of COVID-19 data. Although challenges to sharing review data remain, the efforts already made to introduce organizational policies by research institutions, organizations and funders and the development of sharing infrastructure that stimulate and incentivise both data sharing and use are likely to increase the adoption of such practices.

Authors' contributions

Sze Lin Yoong drafted the initial version of the manuscript. Heidi Turon, Alice Grady, Rebecca Hodder and Luke Wolfenden provided substantial intellectual input. All authors

contributed to critical revisions of the manuscript and approved the final version for submission.

References

- Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. *JAMA* 2016;**315**(12):1238–40.
- World Medical Association. Declaration of Helsinki, ethical principles for medical research involving human subjects. *52 nd WMA General Assembly*, Edinburgh, Scotland. 2000.
- Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *JAMA* 2015;**313**(8):793–4.
- World Health Organisation. *Policy on use and sharing of data collected in Member States by the World Health Organization (WHO) outside the context of public health emergencies (Provisional)*. World Health Organisation. <https://www.who.int/about/policies/publishing/data-policy>. Published 2017 (20 October 2021, date last accessed).
- Organisation for Economic Co-operation and Development. *Data governance: Enhancing access to and sharing of data*. <https://www.oecd.org/sti/economy/enhanced-data-access.htm> (20 October 2021, date last accessed).
- Bill & Melinda Gates Foundation. *Information Sharing Approach*. <https://www.gatesfoundation.org/about/policies-and-resources/information-sharing-approach> (20 October 2021, date last accessed).
- UK Research and Innovation. *Data sharing*. <https://www.ukri.org/councils/mrc/guidance-for-applicants/policies-and-guidance-for-researchers/data-sharing/> (20 October 2021, date last accessed).
- Loder E, Groves T. The BMJ requires data sharing on request for all trials. *BMJ* 2015;**350**(may07 4):h2373.
- Wolfenden L, Grimshaw J, Williams CM, Yoong SL. Time to consider sharing data extracted from trials included in systematic reviews. *Syst Rev* 2016;**5**(1):185.
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2* (updated February 2021). www.training.cochrane.org/handbook (Published 2021; 20 January 2021, date last accessed).
- Page MJ, McKenzie JE, Bossuyt PM *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71.
- 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.
- Saldanha IJ, Smith BT, Ntzani E *et al*. The systematic review data repository (SRDR): descriptive characteristics of publicly available data and opportunities for research. *Syst Rev* 2019;**8**(1):334.
- Yoong SL, Hall A, Stacey F *et al*. Nudge strategies to improve health-care providers' implementation of evidence-based guidelines, policies and practices: a systematic review of trials included within Cochrane systematic reviews. *Implement Sci* 2020;**15**(1):1–30.
- Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;**34**(1):79–87.
- Sajobi TT, Li G, Awosoga O *et al*. A comparison of meta-analytic methods for synthesizing evidence from explanatory and pragmatic trials. *Syst Rev* 2018;**7**(1):19.
- Daly JB, Mackenzie LJ, Freund M *et al*. Interventions by health care professionals who provide routine child health care to reduce tobacco smoke exposure in children: a review and meta-analysis. *JAMA Pediatr* 2016;**170**(2):138–47.
- Rosen LJ, Noach MB, Winickoff JP, Hovell MF. Parental smoking cessation to protect young children: a systematic review and meta-analysis. *Pediatrics* 2012;**129**(1):141–52.
- Nyanchoka L, Tudur-Smith C, Thu VN *et al*. A scoping review describes methods used to identify, prioritize and display gaps in health research. *J Clin Epidemiol* 2019;**109**:99–110.
- Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. *Clin Epidemiol Global Health* 2019;**7**(2):192–8.
- Doubal FN, Ali M, Batty GD *et al*. Big data and data repurposing - using existing data to answer new questions in vascular dementia research. *BMC Neurol* 2017;**17**(1):72.
- Krishnaratne S, Littlecott H, Sell K *et al*. Measures implemented in the school setting to contain the COVID-19 pandemic: a rapid review. *Cochrane Database Syst Rev* 2022;**2022**(2).
- Viswanathan M, Kahwati L, Jahn B *et al*. Universal screening for SARS-CoV-2 infection: a rapid review. *Cochrane Database Syst Rev* 2020;**2020**(9).
- Cochrane. *Cochrane Rapid Reviews*. <https://covidreviews.cochrane.org/>. Published 2022 (20 January 2021, date last accessed).
- Haby MM, Chapman E, Clark R *et al*. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Syst* 2016;**14**(1):83.
- Tricco AC, Antony J, Zarin W *et al*. A scoping review of rapid review methods. *BMC Med* 2015;**13**(1):224.
- van Panhuis WG, Paul P, Emerson C *et al*. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;**14**(1):1144.
- Cochrane Community. *Strategy for Change: 2021-2023*. <https://community.cochrane.org/organizational-info/plans/strategy-change-2021-2023> (20 October 2021, date last accessed).
- Schmidt L, Shokraneh F, Steinhausen K, Adams CE. Introducing RAPTOR: RevMan parsing tool for reviewers. *Syst Rev* 2019;**8**(1):151.
- Kalkman S, Mostert M, Gerlinger C *et al*. Responsible data sharing in international health research: a systematic review of principles and norms. *BMC Med Ethics* 2019;**20**(1):21.
- Cochrane Library. *Terms and conditions for downloading data*. <https://www.cochranelibrary.com/about/data-download> (20 October 2021, date last accessed).
- Hrynaszkiewicz I, Cockerill MJ. *Open by Default: a Proposed Copyright License and Waiver Agreement for Open Access Research and Data in Peer-Reviewed Journals*, Vol. 5. Springer, 2012, 1–12.
- National Institutes of Health. *NIH Data Sharing Policy*. https://grants.nih.gov/grants/policy/data_sharing/. Published 2020 (20 October 2021, date last accessed).
- Council of the European Union. *Council Decision on the Open Data Policy of the Council and the Reuse of Council Documents*. <https://data.europa.eu/doc/document/ST-11563-2017-INIT/en/pdf>. Published 2017 (20 October 2021, date last accessed).

35. The BMJ. *Data Sharing*. <https://authors.bmj.com/policies/data-sharing/> (20 October 2021, date last accessed).
36. Committee on Strategies for Responsible Sharing of Clinical Trial Data, Board on Health Sciences Policy, Institute of Medicine. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington (DC): National Academies Press (US), 2015.
37. Tudur Smith C, Hopkins C, Sydes MR *et al*. How should individual participant data (IPD) from publicly funded clinical trials be shared? *BMC Med* 2015;**13**(1):298.