SYNTHESIS-READY RESEARCH

Ensuring Prevention Science Research is Synthesis-Ready for Immediate and Lasting Scientific Impact

Emily A. Hennessy, University of Connecticut, Institute for Collaboration on Health, Intervention, & Policy, Department of Psychological Sciences, United States (corresponding), emily.hennessy@uconn.edu; https://orcid.org/0000-0002-5146-5823
Rebecca L. Acabchuk, University of Connecticut, Institute for Collaboration on Health, Intervention, & Policy, Department of Psychological Sciences, United States, https://orcid.org/0000-0001-8254-2396
Pieter A. Arnold, Research School of Biology, The Australian National University, Canberra, Australia, https://orcid.org/0000-0002-6158-7752
Adam G. Dunn, Discipline of Biomedical Informatics and Digital Health, School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia. adam.dunn@sydney.edu.au; https://orcid.org/0000-0002-1720-8209
Yong Zhi Foo, Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia. http://orcid.org/0000-0001-7627-2991
Blair T. Johnson, Department of Psychological Sciences, 406 Babbidge Road, Unit 1020, University of Connecticut, Storrs CT 06269-1020 USA, blair.t.johnson@uconn.edu, http://orcid.org/0000-0002-2688-0398
Sonya R. Geange, Research School of Biology, The Australian National University, Canberra, Australia and also Department of Biological Sciences, University of Bergen, Bergen, Norway, https://orcid.org/0000-0001-5344-7234
Neal R. Haddaway, Stockholm Environment Institute (Sweden), Mercator Research Institute on Global Commons and Climate Change (Germany), Africa Centre for Evidence (South Africa) https://orcid.org/0000-0003-3902-2234
Shinichi Nakagawa, Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia https://orcid.org/0000-0002-7765-5182
Witness Mapanga, Non-Communicable Diseases Research Division of the Wits Health Consortium, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. https://orcid.org/0000-0001-8719-3084
Kerrie Mengersen, School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia https://orcid.org/0000-0001-8625-9168
Matthew J. Page, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia, https://orcid.org/0000-0002-4242-7526
Alfredo Sánchez-Tójar, Department of Evolutionary Biology, Bielefeld University, Germany, https://orcid.org/0000-0002-2886-0649
Vivian Welch, Bruyere Research Institute, Ottawa, Canada; School of Epidemiology and Public Health, University of Ottawa, Ottawa Canada; https://orcid.org/0000-0002-5238-7097
Luke A. McGuinness, Department of Population Health Sciences, University of Bristol, United Kingdom, https://orcid.org/0000-0001-8730-9761

Abstract

Synthesis of evidence from the totality of relevant research is essential to inform and improve prevention efforts and policy. Given the large and usually heterogeneous evidence available, reaching a thorough understanding of what works, for whom, and in what contexts, can only be achieved through a systematic and comprehensive synthesis of evidence. Many barriers impede comprehensive evidence synthesis, which leads to uncertainty about the generalizability of intervention effectiveness, including: inaccurate terminology titles/abstracts/keywords (hampering literature search efforts); ambiguous reporting of study methods (resulting in inaccurate assessments of study rigor); and poorly reported participant characteristics, outcomes, and key variables (obstructing the calculation of an overall effect or the examination of effect modifiers). To address these issues and improve the reach of primary studies through their inclusion in evidence syntheses, we provide a set of practical guidelines to help prevention scientists prepare synthesis-ready research. We use a recent mindfulness trial as an empirical example to ground the discussion and demonstrate ways to ensure: (1) primary studies are discoverable; (2) the types of data needed for synthesis are present; and (3) these data are readily synthesizable. We highlight several tools and practices that can aid authors in these efforts, such as creating a repository for each project to host all study-related data files. We also provide step-by-step guidance and software suggestions for standardizing data design and public archiving to facilitate synthesis-ready research.

Ensuring Prevention Science Research is Synthesis-Ready for Immediate and Lasting Scientific Impact

Systematic reviews are vital means of summarizing research to inform policy and practice because they can provide a robust synthesis of available evidence, accounting for risks of bias in individual studies, and because the volume of primary research being published makes it difficult to update policy and practice when new evidence becomes available (Gurevitch et al., 2018; Johnson & Hennessy, 2019; Wilson & Tanner-Smith, 2014).

To conduct a comprehensive evidence synthesis, such as a systematic review and meta-analysis, primary studies related to the research aim must be identifiable and report relevant study and outcome data. The goals of evidence synthesis are often quite different than that of a trial; as a result, trialists may unintentionally omit information that can be crucial for secondary analysis. Unfortunately, the lack of transparency in the reporting of primary studies often makes the work of those conducting evidence syntheses—hereafter labeled *evidence synthesists*—slower and less accurate (Borah et al., 2017; Haddaway & Westgate, 2019; Nakagawa et al., 2020; Nuijten et al., 2016; Wood et al., 2018). Evidence synthesists often need information about measured variables aside from the manuscripts' primary outcome of interest to explore why there are differences (i.e., statistical heterogeneity) between studies examining the same intervention. This information is often missing, which leads to uncertainty about the generalizability of intervention effectiveness across different groups (e.g., race/ethnicity, gender, age). Even when studies report their statistical data and results in detail, there is often still important information missing about other factors that require further attention including author correspondence before they can be accurately synthesized (e.g., description of the intervention, population, or outcome measures).

Given the potential impact of their work on human health and behaviors, prevention scientists should be at the forefront of ensuring primary research is transparent, reproducible and furthermore, we argue, synthesis-ready. *Synthesis-ready research* is our term to describe a study (e.g., randomized controlled trial, RCT), its *meta-data* (i.e., descriptive information about the study's data, including protocols, tools, algorithms, so that one could replicate the study), and *data* (information collected while conducting the study) that can be readily and accurately used in evidence syntheses. Data and metadata may be missing from a primary study for many reasons, including article length restrictions, a lack of detailed reporting standards or awareness of the importance of open data, or a desire to retain control of analyses, amongst others.

Resources are available to enable authors to share detailed data and metadata associated with a study; yet, these resources are often not directly connected to scientific journal practices. Some of these options make it much easier to discover, use, and cite authors' research (for example, in evidence syntheses); substantially increasing the impact and legacy of their work. Thus, our aim here is to draw on resources from across disciplines (Grant et al., 2018; Montgomery et al., 2018; Schulz et al., 2010; Wilkinson et al., 2016) to produce a guide to reporting primary research that is fully open/accessible and synthesis-ready. To illustrate the principles of synthesis-ready research, we draw on an empirical example from a RCT examining mindfulness meditation to improve college student well-being.

1. From Primary Study Research to Evidence Synthesis

The past 10 years have produced advances in primary research study[1] conduct and reporting guidelines, which advocate for reporting enough information for research to be transparent and reproducible and provide specific information on how to do so (Aw et al., 2013;

[1] Primary research can be composed from any type of study design, including experimental (such as randomized controlled trials or controlled trials), observational (such as cohort studies), or mixed methods, to name a few.

Grant et al., 2018; Schulz et al., 2010). Similarly, recent efforts have focused on improving the degree of FAIRness of primary studies, by ensuring research is **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (Figure 1; Wilkinson et al., 2016).[2] By following Open Science (Frankenhuis & Nettle, 2018; Vicente-Saez & Martinez-Fuentes, 2018) recommendations and addressing FAIR principles, prevention scientists can provide synthesis-ready research that improves the translation of research into practice.

Similar to the conduct of primary research, there are best practices for conducting and reporting of evidence syntheses, including the following: problem formulation; systematically identifying and selecting studies for inclusion; coding studies for key features; calculating effect sizes (if synthesizing quantitative data, such as by conducting a meta-analysis); analyzing the data (assess quality and risk of bias, synthesize coded information from studies); interpreting and communicating results; and eventually, updating and re-analyzing data (Online, Figure 1).

2. Potential Advantages of Synthesis-Ready Research

Primary research that is synthesis-ready can benefit the entire discipline as well as primary researchers. Synthesis-ready research can increase the transparency, integrity, and reproducibility of research, ultimately leading to improved scientific evidence (Beugelsdijk et al., 2020). Pre-registration and subsequent data sharing practices are expected to reduce the prevalence of post-hoc hypothesis (HARKing; Forstmeier et al., 2017; Kerr, 1998), and falsifying or misrepresenting data in results (Banks et al., 2019; Miyakawa, 2020). Doing so can reduce duplication of effort and introduce findings to the research community sooner, and ensure that all findings (beneficial and adverse) are reported, results that can prevent costly mistakes (Gupta et al., 2015) and accelerate the impact of prevention scientists (Pasquetto et al., 2019).

[2] Although FAIR principles primarily refer to enhancing the ability of machines to find and use data, these recommendations also enhance data reuse by individuals. Where appropriate, we have adapted the FAIR components to more directly apply to human-readability efforts for the purposes of this manuscript.

Primary researchers who make their work synthesis-ready should also benefit from these efforts. Synthesis-ready research is more likely to be identified and used in evidence synthesis (Christensen et al., 2019; Colavizza et al., 2020; Gerstner et al., 2017). For example, a recent study showed that hosting data in a repository resulted in 25% more citations than merely stating data were available by request (Colavizza et al., 2020). Additionally, depending on the organization of the metadata and the data repository used, the dataset may be cited directly, as may other study materials such as the analytical code, further enhancing the profile of the work and those who conducted it (Crosas, 2013).

Because synthesis-ready research is also more identifiable, there are increased opportunities for future collaborations between primary study researchers, other teams, and evidence synthesists (Popkin, 2019), which potentially improves future resource-sharing and reduces costs. Producing synthesis-ready research can also increase the recognition of an individual study team to an international audience. Many funding sources require data sharing of their grantees and necessitate the inclusion of such plans in grant proposals: Those who prepare and deliver grant-funded research findings with synthesis-ready considerations are considered responsive to funders while potentially attaining these other discussed benefits.

Thus, there are likely many benefits to making research synthesis-ready, despite the slow uptake of openly sharing research data in the social sciences. We will now introduce our empirical example (Section 3) and demonstrate the application of best practices for synthesis-ready research (Sections 4 and 5) to streamline this process for primary study researchers.

3. Introduction to Empirical Example

Our empirical example is a RCT (*N*=140 participants; registered in Clinical Trials, NCT03402009) conducted across two university semesters in the 2018 academic year (Acabchuk

et al, in press). The study evaluated two active treatment conditions to determine what tools best assist university students in developing a personal meditation practice to self-manage symptoms of depression, anxiety, and stress. One group was assigned to the *10% Happier* meditation phone app (labeled the "App Group"; *Ten Percent Happier*, 2020), which provides guided meditations, and the other was assigned to the *10% Happier* meditation phone app in combination with a Muse EEG neurofeedback device (labeled the "Muse Group"; Choose Muse, 2018). The study included a baseline assessment, randomization to one of two groups, an orientation session for both groups, collection of salivary biomarkers, and follow-up assessments on a variety of well-being and behavior outcomes throughout the study (Figure 2).

4. Evidence Synthesis Process Applied to our Empirical Example

Drawing from our empirical example, if evidence synthesists sought to create a systematic review of all types of mindfulness interventions for college students on health-related outcomes (*Problem Formulation*), they would first need to be able to identify all primary research matching their inclusion criteria (*Finding and Selecting Studies*). Clearly delineating the research problem is a foundational step in a review, as it drives the search strategy and ability to accurately identify and synthesize primary study research. Often, evidence synthesists are encouraged to develop research questions that address all the main elements of the studies they are looking to include; many draw on the PICOS framework with its focus on the population(s), intervention(s), comparator(s), outcome(s), and study design(s) elements of a research problem (Stern et al., 2014). Once the relevant literature is identified, evidence synthesists extract data from the included studies (*Code Studies for Relevant Features*) so that the analysis can commence by synthesizing coded data (*Analyze the Systematic Review Database*).

Within this broader process, the types of data and meta-data that need to be reported depend on the task and/or type of evidence synthesis. For example, research methods need to be fully reported to assess risk of bias. If evidence synthesists sought to measure the effectiveness of mindfulness interventions on depression symptoms using a quantitative synthesis, then included primary studies should report an effect size or the statistics necessary to calculate an effect size from their depressive symptom assessment for all participants (e.g., means, standard deviations, and sample sizes for each group). To address relevant potential effect modifiers (e.g., age, year in school, gender, intervention fidelity), any data collected during the primary study on participant characteristics and the study methods (e.g., description of groups, duration of intervention) needs to be transparently reported. Additionally, if the synthesis focuses on participants' perspectives on why interventions are more or less effective, they would need access to any qualitative data collected as part of the study (e.g., interviews, focus groups). Given this variability in the type of questions that could be addressed in a synthesis, it is important for investigators to transparently and comprehensively report a variety of information about their research.[3]

5. How to Make Research Synthesis-Ready: Steps from Study Design to Completion

For many researchers, data sharing only becomes relevant towards the end of a project, when the analyses are completed and the manuscript is ready for submission. Approaching every step of the project with data sharing in mind is useful and often necessary to ensure that data sharing is possible. It has the added benefit of encouraging better overall project management, including clearer workflows and comprehensive documentation of the analytical process. Best practice in data sharing at each stage of a project life cycle include study registration; transparent

[3] Tables 2 and 3 in the appendix present a mapping of critical elements for a study *protocol* using the SPIRIT guidelines (Aw et al., 2013) and for a study *manuscript* using the Consolidated Standards of Reporting Trials for social and psychological interventions (CONSORT-SPI 2018; Grant et al., 2018)) against how study information to be provided based on these guidelines could be used in an evidence synthesis.

workflows; meta-data for identifying authors, results data, published article; and well crafted text to make the reporting easy to find in bibliographic databases (Figure 3).

Study Design

Preparing research for sharing can be a resource-intensive process, and so planning and provisioning for this task is vital to consider when submitting funding applications and considering funders' reporting requirements. Many funders, including the Wellcome Trust, the UK Medical Research Council, and the US National Institutes of Health require data sharing plans as part of grant applications and budget justifications (Gaba et al., 2020; *NIH Data Sharing Policy and Guidance*, 2003; Wellcome Trust Public Health Research Data Forum, 2012).

When conducting research with human participants, planning for data sharing during study design is essential because it ensures appropriate documentation is prepared. That is, any potential ethical concerns with data sharing should be addressed in advance during the review of study materials (i.e., study protocol, advertisements, consent forms) by the governing Institutional Review Board, which ensures that consent documents for participants include the correct language. Then, participants can consent to anonymized data sharing (Ohmann et al., 2017; Thorogood & Knoppers, 2017).

The study protocol is the foundation of the study, detailing all the processes, procedures and materials planned for a research project. Best practice is to pre-register a study protocol that includes primary questions and/or hypotheses, methodology, and planned analyses. With careful attention to developing a comprehensive protocol using a standard structure, this document can increase the synthesis-readiness of the research. The information provided in a study protocol includes important metadata that may not be available in a single published manuscript but that can be used by evidence synthesists to address a variety of research questions, including using it

to assess study quality and risk of bias due to selection of reported results or "outcome switching" (Falk Delgado & Falk Delgado, 2017; Kahan & Jairath, 2018). We recommend that trialists review published guidelines for the best tool for the particular study design in question: The EQUATOR network provides an online resource to compare these tools (https://www.equator-network.org/). The protocol for our mindfulness trial uses the *Standard Protocol Items: Recommendations for Interventional Trials* (SPIRIT) guidelines (Aw et al., 2013) and is archived in a Zenodo repository (DOI: 10.5281/zenodo.4011716).

Study Conduct

To facilitate data sharing and to distribute the workload across the projects' lifecycle, transparent and reproducible data management workflows should be employed throughout the research process, including: (1) defining data file contents; (2) clearly describing variable names, definitions, units and formats; (3) implementing consistent data organization and file structure; (4) performing basic quality assurance to ensure that values fall within expected ranges; and (5) providing documentation including the analytical steps taken (crosswalks), data dictionaries (i.e., codebooks), and metadata (Barton et al., 2010; Strasser, 2012; Yenni et al., 2019). Building in these metadata collection steps into the analysis process makes it easier to create items like crosswalks and data dictionaries to accompany shared research.

Preparing for Study Dissemination

The most labor-intensive stage of making data synthesis-ready occurs during preparation for study dissemination as it involves several steps.

Findability and Accessibility. To address the first two components of FAIR, Findable and Accessible, authors should: (1) ensure all authors have a unique permanent identifier such as an ORCID iD (Open Researcher and Contributor ID; 2009) because institutional emails are often

impermanent, and include this identifier with manuscript and data submission; (2) include a detailed data availability statement with manuscript submission; (3) assign data and metadata a globally unique and persistent identifier (e.g. a DOI), which ensures long-term discoverability; (4) use available data 'tags' in the archiving repository ; (5) use consistent names/usernames across the different journal and archiving platforms; and (6) craft the title and abstract, and select keywords that will optimize the chances of research being identified via the most typical evidence synthesis method of searching for research (i.e., electronic database searches). Open access publishing and/or publishing a preprint of the manuscript will increase the findability of the research; yet, these options may be restricted by available funds and journal policies. We elaborate on some of these points in the sections below.

Data availability statements. Data availability statements (DAS; also known as "Data access statements" or "Availability of materials" sections) increase data findability. A comprehensive DAS provides information about how and where other researchers can access the study data; it is increasingly a required component when submitting a manuscript for publication (Federer et al., 2018; Graf et al. 2020). The DAS should be as explicit as possible, even if some information that is in the method section of a manuscript is repeated in the DAS (Figure 4).[4]

Creating titles, abstracts, and keywords to optimize findability. Carefully crafting titles and abstracts, in addition to choosing appropriate keywords, is essential to help other researchers find your research. As much as possible given word limits, these should describe the condition, exposure, and settings of a study to give readers a quick snapshot of the basics of the research (Grant et al., 2018; Montgomery et al., 2018). These pieces of the submission are often the last to be created, but are the main pieces of information indexed and displayed in electronic databases

[4] In some closed-access journals, DAS are treated as article metadata, meaning that they are accessible even if the full article is behind a paywall, ensuring the relevant data are still open access and locatable

and are used by evidence synthesists to identify relevant research. As a result of their ability to increase the findability of research, we suggest a data-driven approach to crafting these sections. The `litsearchr` package for the R programming environment allows users a systematic method to generate a list of words likely to improve the identifiability of the study manuscript (Grames et al., 2019; R Core Team, 2019).

Using the mindfulness trial, we began with a set of keywords we thought likely to capture a set of relevant published articles from which we could draw the most highly used keywords. We conducted a search using these keywords in PubMed on August 24, 2020 and used `litsearchr` to systematically generate a list of important terms (Table 1). The results suggest that several terms could be strategically incorporated in the title, abstract, and keywords to increase the findability of the trial.

Interoperability and Reusability. To address the third and fourth components of FAIR, Interoperable and Reusable, researchers need to attend to licensing, data file formats, and meta-data when preparing their data for sharing.

Data licenses. Assigning a license to the data being shared is important to govern how the data can be reused, and more importantly, to help ensure that the study team is credited.[5] For our empirical example, we have chosen the CC-BY-4.0 license, which allows the reuse of the data provided the original researchers are credited with the creation of the dataset.

Data file types. Study data can be shared as machine-readable (e.g., unformatted Excel/CSV files and raw DNA microarray sequences) or human-readable formats (e.g., formatted PDF tables, figures, and data contained in text such as "There was evidence of an

[5] Comparing the full suite of licences available is beyond the scope of this article; online resources such as ChooseALicence (https://choosealicense.com/non-software/) and Creative Commons (https://creativecommons.org/licenses/) provide detailed descriptions of the range of licences available for data sharing.

association (Odds ratio: 0.91, 95%CI: 0.85-0.96)"). While new tools are being developed to extract data automatically from human-readable sources for evidence syntheses (Marshall et al., 2020), it is often challenging to import these data into statistical analysis software, reducing interoperability and synthesis-readiness. For our mindfulness trial, all data are shared via Excel files. This includes the mindfulness trial's qualitative data, which were generated from in-person, semi-structured interviews and which were coded to examine adherence and problems with using the apps. To demonstrate synthesis-readiness, the relevant responses for coding are included in the data file rather than the full interview notes.

Meta-data. Providing a README, a plain text file that contains useful information about data or software, can help ensure that data are correctly accessed, interpreted, and reused by other researchers, including by evidence synthesists. The README for our mindfulness trial contains the seven main elements necessary for a comprehensive meta-data file; see Figure 5 for a snapshot of these pieces and visit the Zenodo repository for all files: 10.5281/zenodo.4011716. (1) The README file should include a definition of the primary research aims using the PICOTS framework or some relevant alternative framework that addresses the main components of the research (Haynes, 2006; Johnson & Hennessy, 2019; Stern et al., 2014) which will help evidence synthesists identify and use the research. (2) A short description of all the data files included in the project repository should be included in this file. (3) Signposts to the detailed data dictionaries for each data file should be included; these should contain definitions of all variables, including coded (e.g., "CRP_t2") and interpretable (e.g., "C-Reactive Protein levels at 2 months post-baseline") variable names, the values that each variable is allowed to take, and include missing data indicators and units of measurement. (4) A summary of data processing steps that were used to produce the final analytic datasets from the raw data files should be

provided. (5) A description of each data file's relationship to other data files in the repository, and relevant linkage variables, such as participant ID must be included. (6) The README should have permanent author contact details, including their ORCID. Finally (7), details of the license, including the full name of the license holder, the year, and where the full text of the license can be found should be provided.

Archiving and sharing research data

The two primary considerations when sharing research data include what data to share and where to share it. We provide several potential options within these categories, ordered from best practice (most likely to be synthesis-ready) to good practice (still synthesis-ready but may require additional effort to prepare for synthesis).

Data to share. Sharing individual participant data, i.e., data on each variable collected for each participant, affords evidence synthesists a range of advantages. Access to data on all variables collected, rather than just those used in the primary study's analysis, allows synthesists to ensure that an estimate from a primary study is adjusted for a common set of variables when it is included in a meta-analysis, even if the published effect estimate for that study was not (Riley et al., 2010). Even among individual participant data, different levels of granularity exist. To illustrate the distinction in the level of data shared, for our mindfulness trial, we shared the raw data exported from the survey instrument, which contains each participants' answer to each individual question, and the processed individual participant data, which include scale score summaries (i.e., the total DASS-21 score, which is comprised of 21 individual questions).

It may not be possible to share each individual participant data point, especially when participant confidentiality is at risk; for example, when studying a rare disease or among participants of specialized programs. In these cases, there are several alternative options for

sharing individual participant data. The first is to generate a "synthetic" dataset that mimics the individual-level data by preserving statistical properties and the relationships between variables (Quintana, 2020). Software, for example, the `synthpop` R package, is available to help users generate and explore the usefulness of synthetic datasets (Quintana, 2020, provides a tutorial on using this package). A second option is to remove some individual participant characteristics that might give away participants' identities, such as age, sex, and location of recruitment source, but provide the rest of the variables as individual data points. Finally, if it is not possible to share any individual participant data, then at the very minimum, summary-level data (e.g., means and standard deviations or percentages which describe the distribution of all variables between intervention/exposure groups) should be provided for all variables in the dataset, not solely those thought by the primary study team to be relevant.

Data location. There are several options for sharing research, and a summary of these options, including particular advantages and disadvantages of each, is presented in Table 2.

Non-affiliated repositories. Non-affiliated repositories are commercial or charity repositories that allow researchers to self-archive, and assign a DOI to, their research data. Examples include Dryad, Open Science Framework, and Zenodo (*Dryad*, 2020; *Open Science Framework*, 2020; *Zenodo - Research*. *Shared*., 2020). For our mindfulness trial, the repository was created using GitHub and archived using Zenodo, as GitHub deposition is non-permanent and does not assign a DOI to the uploaded data. Repositories of this type can be embargoed prior to publication to prevent "scooping", and some repositories also provide services to help prevent scooping. For example projects in the Open Science Framework can generate "anonymous" links to remove identifying meta-data as part of double-blind peer-review requirements. The main

limitation of these repositories is that there are often conditions on how much data can be stored; some charge fees for storage.

Institutional repositories. An institutional repository is one hosted by an academic institution, such as the Harvard Dataverse (*Harvard Dataverse*, 2020). An advantage of affiliated repositories is that some provide staff to guide researchers who are new to data sharing. However, these repositories are substantially less well-known, which may limit the findability of data when compared to other options. The terms and conditions of deposition should be reviewed to ensure that authors retain data ownership and copyright over the deposited materials.

Supplemental Files. A common option for sharing research data is as a supplementary file to the published manuscript. However, this method has several major limitations. In the first instance, this approach reduces the findability of the data because datasets shared as supplementary material are grouped under the publication's DOI and are not uniquely identifiable. There is also the potential for supplementary files to become disassociated with their respective publications, for example, during a website redesign. Data shared in this way may also not be accessible; that is, if the published article is behind a paywall, then access to the data attached to it as supplementary files may also be blocked. Sharing data as supplementary files to an open-access publication should be considered the minimum level of data sharing needed to make research synthesis-ready.

6. Potential challenges and future considerations

In this article, we have outlined the benefits of synthesis-ready research and provided guidelines on how to follow best practice and produce synthesis-ready research. Yet, we appreciate that there are many perceived and actual challenges to doing so, including costs to data sharing, lack of incentives, lack of awareness or training, not wanting to "give away" data or

get “scooped” on findings, and the necessity of ensuring participant confidentiality (Evans, 2016; Tennant et al., 2019; Tenopir et al., 2011, 2015; van Panhuis et al., 2014; Walters, 2020).

To address the lack of knowledge and training, we have provided information and workflows here, in addition to signposting to further literature, to help primary researchers plan for the resources needed for making research synthesis-ready even before the project begins. Additionally, many funders have recognized the value of data sharing and will provide funds to do so based on investigator budgets which can offset the tangible (fees for sharing on platforms) and intangible (time spent preparing data for sharing) costs of this process.

Researchers have valid concerns about having their research scooped and yet, several measures can help prevent scooping including pre-registering study analysis plans and depositing preprints while awaiting publication, all core practices of open science. An additional solution is to license data at the point of publication or archival (covered in section 5). Finally, some journals have implemented “scoop-protection” policies for research that has been scooped, a system-level solution we hope more journals will implement to address this challenge (TRANsparency in Scholarly Publishing for Open Scholarship Evolution, 2020).

Ensuring that all types of research data are reused appropriately is a warranted concern (Bishop, 2009; Yardley et al., 2014); yet, evidence suggests that individuals reusing data seek to collaborate with the original data creators to harness their extensive knowledge about the data, rather than using it to scoop the data or to disprove the original publication’s findings, resulting in improved and more accurate data reuse and increased opportunities for publication (Pasquetto et al., 2019). As well, providing clearly detailed documentation in the data and metadata files (discussed in Section 5) should reduce inappropriate use of shared data. Finally, we have also

provided several different options for maintaining the confidentiality of participants while attending to synthesis-readiness (Section 5).

Several advances in computational abilities may help to increase the future of preparing synthesis-ready research. Federated learning or analysis enable data from different studies to be analysed in a combined statistical model, while retaining the confidentiality and privacy requirements of the separate data custodians (Geleijnse et al., 2020). It involves using an iterative process of model building from all available data, while sensitive data is stored in their respective secure institutions. To date, the application of this method has been limited, but it is an appealing future direction for evidence syntheses and primary studies.

7. Conclusions

Despite the importance of making research available and synthesis-ready, not all scientific disciplines have embraced data/metadata sharing and transparency at the same rate (Graf et al., 2020; Hardwicke et al., 2020; Sholler et al., 2019; Vasilevsky et al., 2017; Whitlock, 2011). In this paper, we argue that if researchers attended to sharing research in a way to make it synthesis-ready, they and the field of prevention science should see a variety of benefits. In addition, we present several methods to reduce, or at least more easily manage, the tasks of data sharing and to mitigate the perceived challenges we detail in this manuscript. Our hope is that we will soon see increased attention to practices that ensure the products of primary studies are directly fed into evidence syntheses. This will reduce the delay between implementing effective primary study approaches and changing practice (Nakagawa et al., 2020), thereby improving the scientific process and products for future generations.

Compliance with Ethical Standards

Acknowledgements

The paper was initially developed through the discussion stream at the 2019 Evidence Synthesis Hackathon (www.eshackathon.org), an inter- and trans-disciplinary meeting of evidence synthesists and those who develop software tools for evidence synthesis. Project administration (EAH; LAM), Roles/Writing - original draft (EAH; LAM), Visualization (EAH; LAM); Data curation (RLA; LAM); Resources (LAM); Funding acquisition (NRH); Investigation (EAH; LAM; RLA), Methodology (EAH; LAM); Software (EAH; LAM); Conceptualization (All authors), Writing - review & editing (All authors)

Funding

The Evidence Synthesis Hackathon 2019 was funded by the Fenner School of Environment & Society at the Australian National University, the University of New South Wales, and the Faculty of Humanities at the University of Johannesburg. EAH has support from NIAAA (K01 AA028536-01). MJP has support from an Australian Research Council Discovery Early Career Researcher Award (DE200101618). LAM has support by an NIHR Doctoral Research Fellowship (DRF-2018-11-ST2-048). The views expressed in this article are those of the authors and do not necessarily represent those of the NIAAA, NHS, the NIHR, or the Department of Health and Social Care.

Disclosure of Potential Conflicts of Interest: None identified.

Ethics Approval: Not applicable

Informed Consent: Not applicable

References

Abbott, S., Christensen, H., Welton, N. J., & Brooks-Pollock, E. (2019). Estimating the effect of the 2005 change in BCG policy in England: A retrospective cohort study, 2000 to 2015. *Eurosurveillance*, *24*(49), 1900220. https://doi.org/10/gg9nqk

Chan, A. W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., ... & Doré, C. J. (2013). *SPIRIT 2013 statement: Defining standard protocol items for clinical trials*. Annals of Internal Medicine; Ann Intern Med. https://doi.org/10.7326/0003-4819-158-3-201302050-00583

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 Questions About Open Science Practices. *Journal of Business and Psychology*, *34*(3), 257–270. https://doi.org/10/gdj7sx

Barton, C., Smith, R., & Weaver, R. (2010). Data Practices, Policy, and Rewards in the Information Era Demand a New Paradigm. *Data Science Journal*, *9*, IGY95–IGY99. https://doi.org/10/d7sjcz

Beugelsdijk, S., van Witteloostuijn, A., & Meyer, K. E. (2020). A new approach to data access and research transparency (DART). *Journal of International Business Studies*, *51*(6), 887–905. https://doi.org/10/gg57tv

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*(2), e012545. https://doi.org/10.1136/bmjopen-2016-012545

Choose Muse. (2018, June 25). *A Deep Dive Into Brainwaves: Brainwave Frequencies Explained*. Muse. https://choosemuse.com/blog/a-deep-dive-into-brainwaves-brainwave-

frequencies-explained-2/

Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., & Rose, A. K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLOS ONE*, *14*(12), e0225883. https://doi.org/10/gghfvf

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, *15*(4), e0230416. https://doi.org/10/ggtcrb

Crosas, M. (2013, October 30). *Joint Declaration of Data Citation Principles—FINAL*. FORCE11. https://www.force11.org/datacitationprinciples

Dryad. (2020). https://datadryad.org/stash/our_mission

Evans, S. R. (2016). Gauging the Purported Costs of Public Data Archiving for Long-Term Population Studies. *PLOS Biology*, *14*(4), e1002432. https://doi.org/10/f8kf9w

Falk Delgado, A., & Falk Delgado, A. (2017). Outcome switching in randomized controlled oncology trials reporting on surrogate endpoints: A cross-sectional analysis. *Scientific Reports*, *7*. https://doi.org/10/gbvdzn

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y.-L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS ONE*, *13*(5), e0194768. https://doi.org/10/gdhb7j

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10/gdf37n

Frankenhuis, W. E., & Nettle, D. (2018). Open Science Is Liberating and Can Foster Creativity. *Perspectives on Psychological Science*, *13*(4), 439–447.

https://doi.org/10.1177/1745691618767878

Gaba, J. F., Siebert, M., Dupuy, A., Moher, D., & Naudet, F. (2020). Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PLOS ONE*, *15*(8), e0237464. https://doi.org/10/gg87d2

Geleijnse, G., Chiang, R. C.-J., Sieswerda, M., Schuurman, M., Lee, K. C., van Soest, J., Dekker, A., Lee, W.-C., & Verbeek, X. A. A. M. (2020). *Prognostic Factors for Survival in Patients with Oral Cavity Cancer: A Comparison of the Netherlands and Taiwan Using Privacy-Preserving Federated Analyses* (SSRN Scholarly Paper ID 3539690). Social Science Research Network. https://doi.org/10.2139/ssrn.3539690

Gerstner, K., Moreno-Mateos, D., Gurevitch, J., Beckmann, M., Kambach, S., Jones, H. P., & Seppelt, R. (2017). Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods in Ecology and Evolution*, *8*(6), 777–784. https://doi.org/10.1111/2041-210X.12758

Graf, C., Flanagan, D., Wylie, L., & Silver, D. (2020). The Open Data Challenge: An Analysis of 124,000 Data Availability Statements and an Ironic Lesson about Data Management Plans. *Data Intelligence*, 1–15. https://doi.org/10/gg87d3

Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, *10*(10), 1645–1654. https://doi.org/10/ggd7g4

Grant, S., Mayo-Wilson, E., Montgomery, P., Macdonald, G., Michie, S., Hopewell, S., Moher, D., Aber, J. L., Altman, D., Bhui, K., Booth, A., Clark, D., Craig, P., Eisner, M., Fraser, M. W., Gardner, F., Grant, S., Hedges, L., Hollon, S., … on behalf of the CONSORT-SPI

Group. (2018). CONSORT-SPI 2018 Explanation and Elaboration: Guidance for reporting social and psychological intervention trials. *Trials*, *19*(1), 406. https://doi.org/10/gd2p8t

Gupta, Y. K., Meenu, M., & Mohan, P. (2015). The Tamiflu fiasco and lessons learnt. *Indian Journal of Pharmacology*, *47*(1), 11–16. https://doi.org/10/gg898k

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175–182. https://doi.org/10.1038/nature25753

Haddaway, N. R., & Westgate, M. J. (2019). Predicting the time needed for environmental systematic reviews and systematic maps. *Conservation Biology*, *33*(2), 434–443. https://doi.org/10/ggnn3p

Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. A. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, *7*(2), 190806. https://doi.org/10/ggm5vg

Harvard Dataverse. (n.d.). Retrieved July 31, 2020, from https://dataverse.harvard.edu/

Haynes, R. B. (2006). Forming research questions. *Journal of Clinical Epidemiology*, *59*(9), 881–886. https://doi.org/10.1016/j.jclinepi.2006.06.006

Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine*, *233*, 237–251. https://doi.org/10/ggxrwv

Kahan, B. C., & Jairath, V. (2018). Outcome pre-specification requires sufficient detail to guard against outcome switching in clinical trials: A case study. *Trials*, *19*(1), 265. https://doi.org/10/gdjsqg

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and social psychology review*, *2*(3), 196-217.

Marshall, I. J., Johnson, B. T., Wang, Z., Rajasekaran, S., & Wallace, B. C. (2020). Semi-Automated evidence synthesis in health psychology: Current methods and future prospects. *Health Psychology Review*, *14*(1), 145–158. https://doi.org/10.1080/17437199.2020.1716198

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, *13*(1), 24. https://doi.org/10.1186/s13041-020-0552-2

Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., Moher, D., Lawrence Aber, J., Altman, D., Bhui, K., Booth, A., Clark, D., Craig, P., Eisner, M., Fraser, M. W., Gardner, F., Grant, S., Hedges, L., Hollon, S., … on behalf of the CONSORT-SPI Group. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 Extension. *Trials*, *19*(1), 407. https://doi.org/10.1186/s13063-018-2733-1

Nakagawa, S., Dunn, A. G., Lagisz, M., Bannach-Brown, A., Grames, E. M., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Westgate, M. J., Arnold, P. A., Barrow, S., Bethel, A., Cooper, E., Foo, Y. Z., Geange, S. R., Hennessy, E., Mapanga, W., Mengersen, K., Munera, C., … Evidence Synthesis Hackathon 2019 Participants. (2020). A new ecosystem for evidence synthesis. *Nature Ecology & Evolution*, *4*(4), 498–501. https://doi.org/10.1038/s41559-020-1153-2

NIH data sharing policy and guidance. (2003). National Institutes of H ealth. https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M.

(2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. https://doi.org/10/f9pdjm

Ohmann, C., Banzi, R., Canham, S., Battaglia, S., Matei, M., Ariyo, C., Becnel, L., Bierer, B., Bowers, S., Clivio, L., Dias, M., Druml, C., Faure, H., Fenner, M., Galvez, J., Ghersi, D., Gluud, C., Groves, T., Houston, P., … Demotes-Mainard, J. (2017). Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open*, *7*(12), e018647. https://doi.org/10/gcpj3x

Open Science Framework. (2020). https://osf.io/

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10/ggc9n3

Pescosolido, B., Lee, B., & Kafadar, K. (2020). *Replication Data for: Socio-demographic Similarity at the Contextual Level Alters Individual Risk for Completed Suicide* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/35IV23

Popkin, G. (2019). Data sharing and how it can benefit your scientific career. *Nature*, *569*(7756), 445–447. https://doi.org/10/gf2v8j

Quintana, D. S. (2020). A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *ELife*, *9*, e53275. https://doi.org/10/gg72dz

R Core Team. (2019). *R: A language and environment for statistical computing*. https://www.R-project.org/

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, *340*. https://doi.org/10/fvk3mc

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010

statement: Updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical Research Ed.)*, *340*, c332. https://doi.org/10.1136/bmj.c332

Sholler, D., Ram, K., Boettiger, C., & Katz, D. S. (2019). Enforcing public data archiving policies in academic publishing: A study of ecology journals. *Big Data & Society*, *6*(1), 2053951719836258. https://doi.org/10.1177/2053951719836258

Slep, A. M. S., Heyman, R. E., Lorber, M. F., Baucom, K. J. W., & Linkh, D. J. (2020). Evaluating the Effectiveness of NORTH STAR: A Community-Based Framework to Reduce Adult Substance Misuse, Intimate Partner Violence, Child Abuse, Suicidality, and Cumulative Risk. *Prevention Science*. https://doi.org/10/gg898m

Stern, C., Jordan, Z., & McArthur, A. (2014). Developing the Review Question and Inclusion Criteria. *AJN The American Journal of Nursing*, *114*(4), 53–56. https://doi.org/10/ggsp4z

Strasser, C. (2012). *Primer on Data Management: What you always wanted to know*. https://doi.org/10/gg6rgg

Ten Percent Happier: Mindfulness Meditation Courses. (2020). Ten Percent Happier. https://www.tenpercent.com

Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B., Ross-Hellauer, T., Sattler, S., Thacker, P. D., & Vanholsbeeck, M. (2019). Ten Hot Topics around Scholarly Publishing. *Publications*, *7*(2), 34. https://doi.org/10/gf4gvx

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, *6*(6), e21101. https://doi.org/10/fgppw4

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett,

K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, *10*(8), e0134826. https://doi.org/10/8fw

Thorogood, A., & Knoppers, B. M. (2017). Can research ethics committees enable clinical trial data sharing? *Ethics, Medicine and Public Health*, *3*(1), 56–63. https://doi.org/10/gg6rgh

Tomczyk, S., Schomerus, G., Stolzenburg, S., Muehlan, H., & Schmidt, S. (2020). Ready, Willing and Able? An Investigation of the Theory of Planned Behaviour in Help-Seeking for a Community Sample with Current Untreated Depressive Symptoms. *Prevention Science*, *21*(6), 749–760. https://doi.org/10/gg898n

TRANsparency in Scholarly Publishing for Open Scholarship Evolution. (2020). *Transpose Database: A database of journal policies on peer review, co-reviewing, and preprinting*. https://transpose-publishing.github.io/#/

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, *14*(1), 1144. https://doi.org/10/f6tgv2

Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ*, *5*. https://doi.org/10/ggb7z3

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

Walters, W. H. (2020). Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights*, *33*(1), 18. https://doi.org/10/gg4rhz

Wellcome Trust Public Health Research Data Forum. (2012). *Summary of Funders' data sharing*

policies. https://wellcome.ac.uk/sites/default/files/summary-of-phrdf-funder-data-sharing-policies.pdf

Whitlock, M. C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, *26*(2), 61–65.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10/bdd4

Wilson, S. J., & Tanner-Smith, E. E. (2014). Meta-analysis in Prevention Science. In Z. Sloboda & H. Petras (Eds.), *Defining Prevention Science* (pp. 431–452). Springer US. https://doi.org/10.1007/978-1-4899-7424-2_19

Wood, B. D. K., Müller, R., & Brown, A. N. (2018). Push button replication: Is impact evaluation evidence for international development verifiable? *PLOS ONE*, *13*(12), e0209416. https://doi.org/10/ggdmth

Yenni, G. M., Christensen, E. M., Bledsoe, E. K., Supp, S. R., Diaz, R. M., White, E. P., & Ernest, S. K. M. (2019). Developing a modern data workflow for regularly updated data. *PLOS Biology*, *17*(1), e3000125. https://doi.org/10/c82f

Zenodo—Research. Shared. (2020). https://about.zenodo.org/

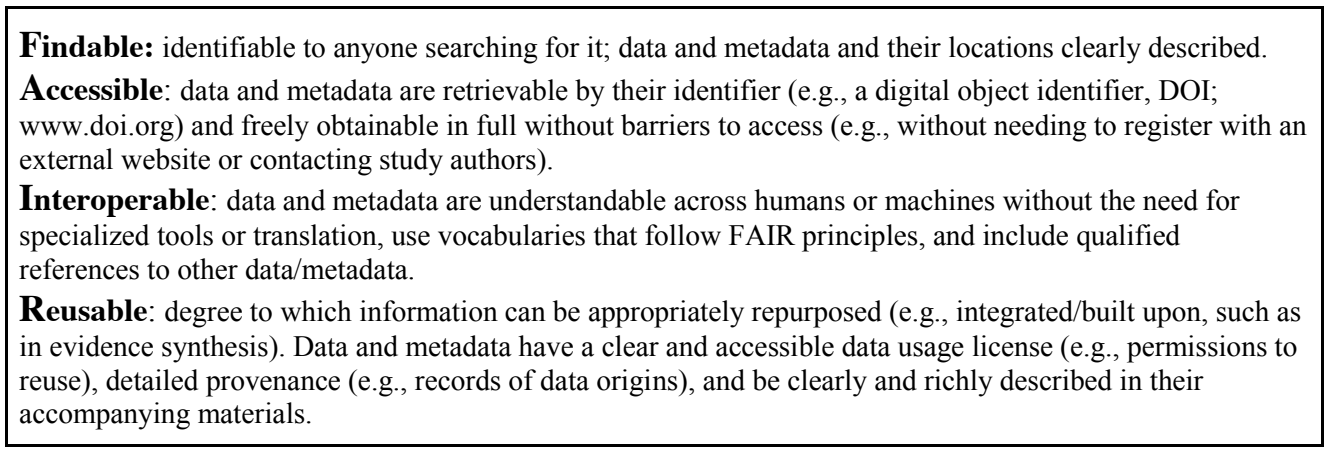


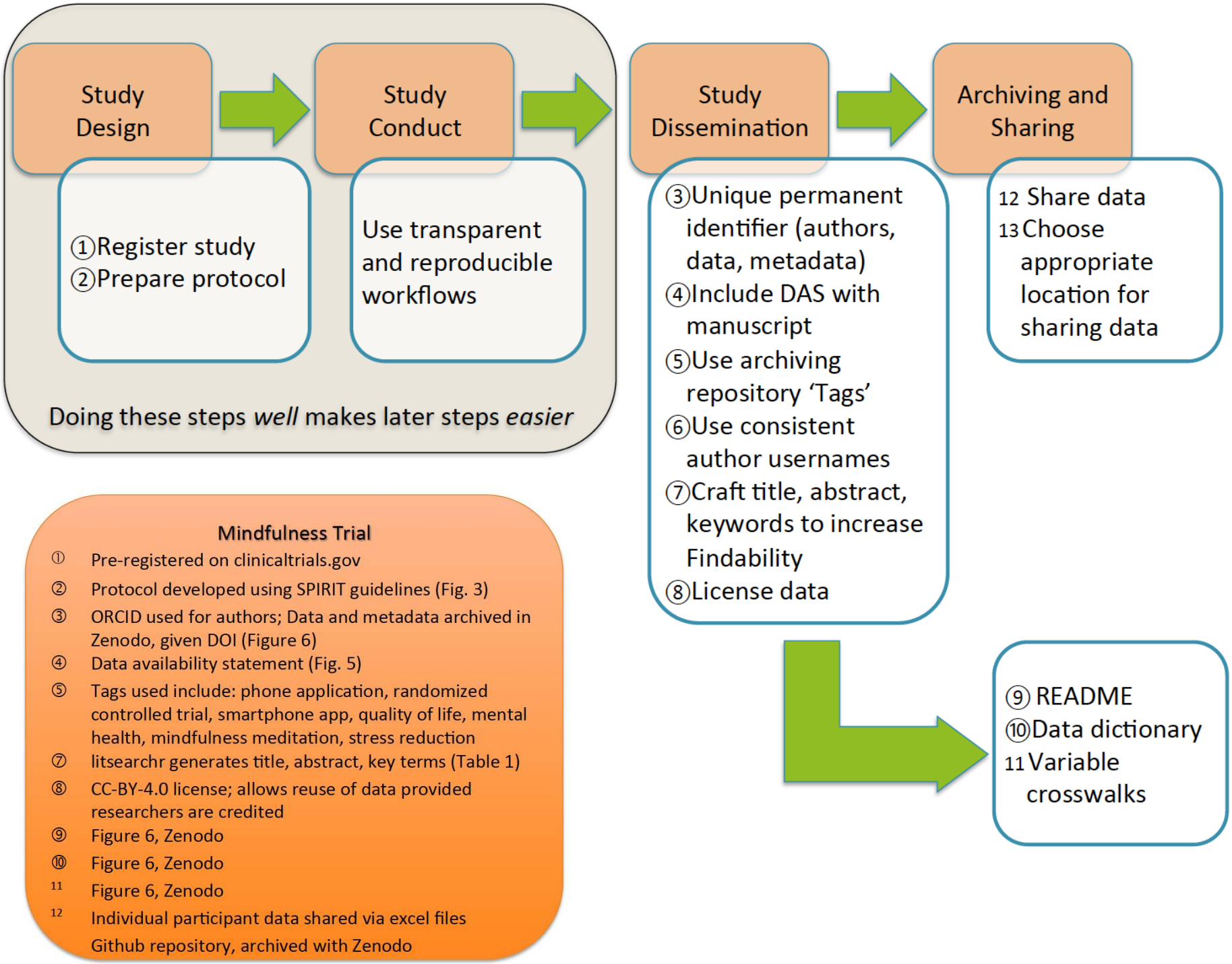
Figure 1. FAIR principles (Wilkinson et al., 2016)

Figure

		STUDY PERIOD				
	Enrolment	Allocation	Post-allocation			Close-out
TIMEPOINT**	$-t_1$	0	*1 month*	*2 months*	*7 months*	*13 months*
ENROLMENT:						
Eligibility screen	X					
Informed consent	X					
1 Hour Orientation	X					
Allocation		X				
INTERVENTIONS:						
App group		→				
Muse group		→				
ASSESSMENTS:						
Demographics, medical history, motivation	X					
Time calm (%) and bird scores during meditation (5 minutes) [a]		X	X			
Salivary biomarkers	X (n = 20)		X (n = 20)			
Meditation tool use; adverse experiences (every 48 hours)		X	X			
Distress, anxiety, stress, depressive symptoms, resilience, sleep, eating, mindfulness, emotion regulation, interoception, decentering, self-esteem, personality	X		X			
Satisfaction			X			
Meditation practice		X	X	X	X	X

Figure 2. Overview of Mindfulness Trial using the SPIRIT template of recommended content for the schedule of enrolment, interventions, and assessments (Aw et al., 2013)
[a] $n = 52$, Wave 1 only

Figure 3. Best Practices in Generating Synthesis-Ready Research applied to Mindfulness Trial



Both the full raw dataset and the finalized analytical dataset used in this analysis are available from Zenodo (DOI: 10.5281/zenodo.4011716), along with comprehensive data dictionaries and crosswalks.

Figure 4. Data availability statement for Mindfulness Trial

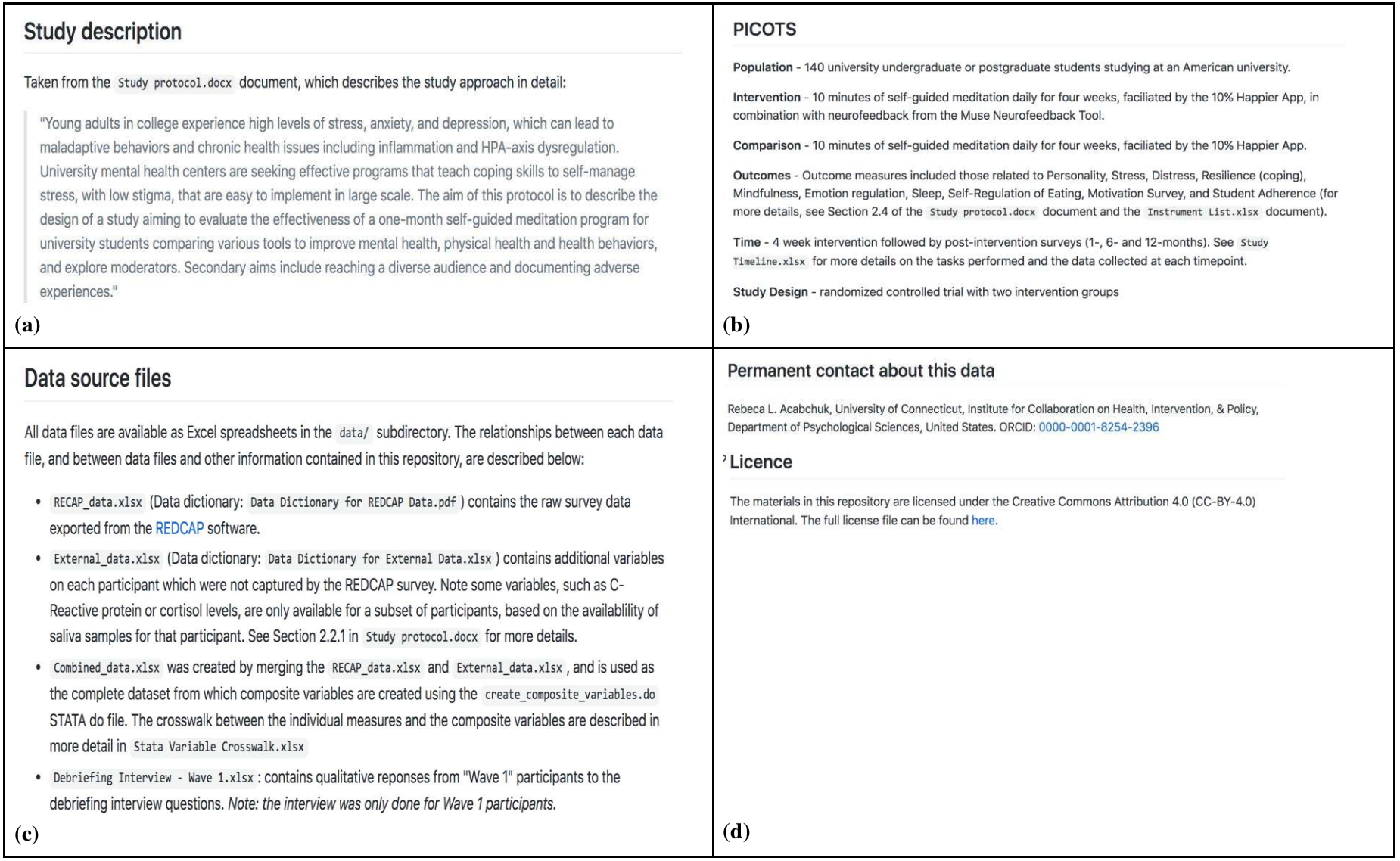


Figure 5. Mindfulness trial metadata included in the README file. Panels (a) and (b) provide the study description according to the PICOTS framework; (c) provides data source files, locations, and relationships between the files; (d) provides permanent author information and licensing information.

Table 1. Demonstration of use of `litsearchr` to produce key terms

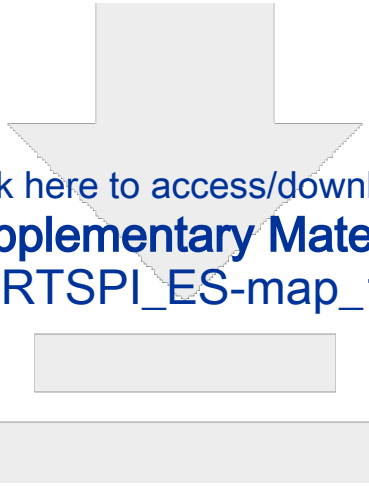
Database	Search Strategy	Top 50 Keywords, ranked from greatest to least important		
PubMed (373 results)	(((((university student[Title/Abstract]) OR (college student[Title/Abstract])) AND (mindfulness[Title/Abstract])) OR (meditation[Title/Abstract])) AND (apps[Title/Abstract])) OR (phone application[Title/Abstract])	phone application mobile phone mobile phone application controlled trial mobile app trial registration randomized control randomized controlled randomized controlled trial control group trial background mobile application smartphone app smart phone quality of life controlled trial background smart phone application mental health	primary outcome pilot study intervention group mobile health secondary outcome health intervention phone applications secondary outcomes study protocol health care clinical trial randomized controlled trial background randomised controlled trial randomised controlled health service physical activity data collection	mobile phones study background smartphone application mobile phone applications health problem study aimed participants completed based intervention health outcomes focus group outcome measure health application outcome measures public health outcomes include
PubMed (1589 results)	(((((((university[Title/Abstract]) OR (college[Title/Abstract])) AND (mindfulness[Title/Abstract])) OR (meditation[Title/Abstract])) AND (treatment[Title/Abstract])) OR (intervention[Title/Abstract])) OR (program[Title/Abstract])) AND (apps[Title/Abstract])) OR (phone application[Title/Abstract])	controlled trial randomized control mobile app randomized controlled randomized controlled trial mobile phone trial registration smartphone app phone application mobile health health app mobile phone app physical activity mobile apps primary outcome trial background control group health intervention	clinical trial health care systematic review controlled trial background based intervention controlled trials health apps intervention group study background secondary outcome mental health randomized controlled trial background smartphone apps health interventions clinical trials quality of life randomised control	mobile phone application randomised controlled outcome measure randomised controlled trial mobile application study aimed behavior change secondary outcomes data collection smartphone application review background mhealth app health outcome phone applications health outcomes
PubMed (4664 results)	((((((eeg[Title/Abstract]) OR (neurofeedback[Title/Abstract])) AND (RCT[Title/Abstract])) OR (randomized controlled trial[Title/Abstract])) AND (mental health[Title/Abstract])) AND (mindfulness[Title/Abstract])) OR (meditation[Title/Abstra	controlled trial randomized control mindfulness meditation randomized controlled control group randomized controlled trial quality of life stress reduction mental health meditation practice based intervention outcome measure based stress based stress reduction	mindfulness-based stress reduction pilot study systematic review significant improvement randomly assigned perceived stress based interventions significant difference mindfulness training mindfulness-based intervention meditation training trial registration	secondary outcome month follow depressive symptom results suggest heart rate trial background active control depressive symptoms mindfulness-based interventions transcendental meditation intervention group mindfulness practice

	ct])	controlled trials primary outcome outcome measures clinical trial	statistically significant present study blood pressure randomized controlled trials	findings suggest showed significant significant reduction significant differences

Note. For the full instructions, including R syntax and our citation files, see https://github.com/mcguinlu/prevsci-data-sharing-exemplar/tree/master/keyword_generation

Table 2. Comparison of the options for sharing data

Location	Examples	Advantages	Disadvantages	Examples
Non-affiliated repositories	Zenodo; Dryad; Figshare; Open Science Framework	Free (except Dryad). Self-archiving. DOI assigned. Long-term storage. Can blind repositories for double-blind peer-review (Anonymous Github; OSF "Anonymous" Links)	Some have a Data Publishing Charge (e.g., Dryad charges $120). No review of files uploaded.	Repository for Abbott et al., 2019, contains raw dataset, analysis dataset, cleaning code, analysis code. Available from 10.5281/zenodo.3529382
Institutional repositories	data.bris; Harvard Dataverse	Supported by institution. DOI assigned. Often can get help from staff when uploading data.	Not as well known as other repositories - limits Findability.	Pescosolido et al., 2020, contains scripts and data to replicate all figures/analyses found in the original manuscript.
Supplemental files	Any file (.txt, .xlsx, .doc, .pdf) with data and attached to the online version of the published article	Likely relatively less work to prepare compared to the other options	No DOI (for data). Can be closed-access. Potential for data to be lost. Least findable of the other options.	Slep et al., 2020, contains mPLUS code and data Tomczyk et al., 2020, open access, includes data in manuscript and supplemental files



Click here to access/download

Supplementary Material

SPIRIT-CONSORTSPI_ES-map_15sep2020.docx