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

Emily A. Hennessy,

Recovery Research Institute, Center for Addiction Medicine, Massachusetts General Hospital & Harvard Medical School, Boston, MA USA

Rebecca L. Acabchuk,

University of Connecticut, Institute for Collaboration on Health, Intervention, & Policy, Department of Psychological Sciences, United States

Pieter A. Arnold,

Research School of Biology, The Australian National University, Canberra, Australia

Adam G. Dunn,

The University of Sydney, Biomedical Informatics and Digital Health, School of Medical Sciences, Faculty of Medicine and Health, Sydney, Australia.

Yong Zhi Foo,

Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia.

Blair T. Johnson,

Department of Psychological Sciences, 406 Babbidge Road, Unit 1020, University of Connecticut, Storrs CT 06269-1020 USA

Sonya R. Geange,

Research School of Biology, The Australian National University, Canberra, Australia and also Department of Biological Sciences, University of Bergen, Bergen, Norway

Neal R. Haddaway,

Stockholm Environment Institute (Sweden), Mercator Research Institute on Global Commons and Climate Change (Germany), Africa Centre for Evidence (South Africa)

Shinichi Nakagawa,

Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

Witness Mapanga,

Non-Communicable Diseases Research Division of the Wits Health Consortium, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

(corresponding), ehennessy@mgh.harvard.edu.

Kerrie Mengersen,

School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

Matthew J. Page,

School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Alfredo Sánchez-Tójar,

Department of Evolutionary Biology, Bielefeld University, Germany

Vivian Welch,

Bruyere Research Institute, Ottawa, Canada; School of Epidemiology and Public Health, University of Ottawa, Ottawa Canada

Luke A. McGuinness

Department of Population Health Sciences, University of Bristol, United Kingdom

Abstract

When seeking to inform and improve prevention efforts and policy, it is important to be able to robustly synthesize all available evidence. But evidence sources are often large and heterogeneous, so 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 titles/abstracts/keywords terminology (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 using a data-driven approach for crafting titles, abstracts, and keywords or by 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.

Systematic reviews are an important tool for informing policy and practice. When done well, they provide a robust synthesis of available evidence, account for risks of bias in individual studies, and address problems associated with the volume of primary research being published (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. Those who conduct evidence syntheses—hereafter referred to as *evidence synthesists* — often need information about study processes and measured variables aside from a 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) and can make these efforts slower and less accurate (Borah et al., 2017; Haddaway & Westgate, 2019; Nakagawa et al., 2020; Nuijten et al., 2016). 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 before they can be accurately synthesized (e.g., description of the intervention, population, or outcome measures). Data and meta-data 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.

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), quasi-experimental study (QED), or observational study), 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.

Resources are available to enable authors to share detailed data and meta-data 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; substantially increasing the impact and legacy of their work. Thus, our aim was 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 open, accessible, and synthesis-ready. To illustrate synthesis-ready research principles and streamline this process for prevention scientists, we briefly present an empirical example from a mindfulness meditation RCT to improve college student well-being (Section 3), followed by a demonstration of the application of best practices for synthesis-ready research to this example (Sections 4 and 5).

1. From Primary Study Research to Evidence Synthesis

Primary study conduct and reporting guidelines now advocate for reporting enough information for research to be transparent and reproducible and also provide specific information on how to do so (Aw et al., 2013; Grant et al., 2018; Schulz et al., 2010). Similarly, efforts have focused on improving the degree of FAIRness of primary studies, by making research **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (Wilkinson et al., 2016).[1]

Findable indicates research is identifiable to anyone searching for it; data and meta-data and their locations clearly described. Data and meta-data that are retrievable by their identifier

---

[1]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.

(e.g., a digital object identifier, DOI) and are freely obtainable in full without barriers to access (e.g., without needing to register with an external website or contacting study authors) are considered **A**ccessible. Data and meta-data that are understandable across humans or machines without the need for specialized tools or translation and include linkage to their relevant data/meta-data are **I**nteroperable. Finally, **R**esuable refers to the degree to which information can be appropriately repurposed (e.g., integrated/built upon). Thus, research data and meta-data should have a data usage license, detailed records of data origins, and be clearly and richly described in their accompanying materials. By following Open Science recommendations and addressing FAIR principles (Frankenhuis & Nettle, 2018; Vicente-Saez & Martinez-Fuentes, 2018), prevention scientists can provide synthesis-ready research that improves the translation of research into practice.

Evidence syntheses rely on the data reported in research studies and by the researchers themselves. 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 (Online, Figure 1). The content and the comprehensiveness of the review (if the review authors follow best practice guidelines), is a direct result of the available information reported in primary studies.

2. Potential Advantages of Synthesis-Ready Research

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 should 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 and accelerate the impact of prevention scientists (Gupta et al., 2015; Pasquetto et al., 2019). Indeed, synthesis-ready data can be used repeatedly and for multiple purposes, ultimately informing more research questions: thus it is more cost-effective when compared to data that is only used by the team that collected it.

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; 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, Hrynaszkiewicz, Staden, Whitaker, & McGillivray, 2020). Additionally, depending on the organization of the meta-data 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 will be increased opportunities for future collaborations (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 (e.g., Wellcome Trust, the UK Medical Research Council, and the US National Institutes of Health: Gaba et al., 2020; NIH Data Sharing Policy and Guidance, 2003; Wellcome Trust Public Health Research Data Forum, 2012). Thus, researchers who prepare and deliver grant-funded research findings with synthesis-ready considerations are considered responsive to funders while potentially attaining these other discussed benefits.

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, Simon, Low, Brisson, & Johnson, 2021). 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 survey assessments on a variety of well-being and behavior outcomes throughout the study (Study Timeline: Online, Figure 2).

4. Evidence Synthesis Process Applied to our Empirical Example

Drawing from our empirical example, if one sought to conduct 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. Research questions in an evidence synthesis should address all the main elements of the studies to be included; 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, data are extracted from the 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 in a primary study 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 the review aimed to measure the effectiveness of mindfulness interventions on depression symptoms using a meta-analysis, then primary studies should report the data 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.

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

Approaching every step of a research project with data sharing in mind is useful and often necessary to ensure that data sharing is possible. This approach has the added benefit of encouraging better overall project management, including clearer workflows and comprehensive documentation of the analytical process. We now illustrate each step of these suggestions from study design to sharing with examples from our mindfulness trial (Figure 1).

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. When conducting research with human participants, advance planning for data sharing is essential because it ensures appropriate documentation is prepared. Any potential ethical concerns with data sharing should be addressed during the review of study materials by the governing Institutional Review Board, which ensures that participant consent documents include the correct language to ensure participants can consent to anonymized data sharing (Ohmann et al., 2017; Thorogood & Knoppers, 2017). The Inter-university Consortium for Political and Social Research (ICPSR) recommends carefully constructing these sections and provides sample text to use to ensure that participants are informed about the potential use of their anonymized data in future research (ICPSR, 2010).

The study protocol is the foundation of the study, detailing all the processes, procedures and materials planned for a research project. Best practice for primary studies and secondary data analysis is to 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 meta-data 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 authors review published guidelines for the best tool for the study design in question: The EQUATOR network provides an online resource to compare these tools (https://www.equator-network.org/). Online Tables 1 and 2 present a mapping of critical elements for a study *protocol* (using the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT); Aw et al., 2013) and for a study *manuscript* (using the Consolidated Standards of Reporting Trials for social and psychological interventions (CONSORT); Grant et al., 2018)) and indicates how study information provided following these guidelines could be used in an evidence synthesis.

The protocol for our mindfulness trial uses the 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 meta-data (Barton et al., 2010; Strasser, 2012; Yenni et al., 2019). Building in these meta-data collection steps into the analysis process makes it easier to create items like variable crosswalks and data dictionaries to accompany shared research.

Study Dissemination

The most labor-intensive stage of making data synthesis-ready within a FAIR perspective occurs during preparation for study dissemination as it involves several steps. To ensure their research is *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 meta-data a globally unique and persistent identifier (e.g., 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). To ensure their data is also *Interoperable* and *Reusable*, researchers need to attend to licensing, data file structures and formats, and detailed meta-data when preparing their data for sharing.

Data Availability Statements to Ensure Findability.—A data availability statement (DAS; also known as "Data access statements" or "Availability of materials") provides information about how and where other researchers can access the study data and is often required when submitting a manuscript for publication (Federer et al., 2018; Graf et al. 2020). In some closed-access journals, DAS are treated as article meta-data, so they are

accessible even if the article is behind a paywall, ensuring the data are findable. Thus, the DAS should be as explicit as possible, even if some information that is in the manuscript is repeated in the DAS.

The DAS for our mindfulness trial example states: "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."

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 (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 and are used to identify relevant research for an evidence synthesis. 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 offers a systematic approach to generate a list of words likely to improve the identifiability of the study manuscript (Grames et al., 2019; R Core Team, 2019).

Using our mindfulness trial example, 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. The results suggest that several terms could be strategically incorporated in the title, abstract, and keywords to increase the findability of the trial (Online, Table 3 provides further details).

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.[2]

For our mindfulness trial example, we chose 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 Structures and Formats.—Sharing individual participant data, i.e., data on each variable collected for each participant, affords those wishing to reuse or synthesize these data 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.

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

---

[2]Comparing the full suite of licenses 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 licenses available for data sharing.

participants of specialized programs. In these cases, there are several 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. Software, for example, the synthpop R package, can help users generate a synthetic dataset (Quintana, 2020, provides a tutorial on using this package). A second option is to remove some individual 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 included in the paper and if possible, the dataset, not solely those thought by the primary study team to be relevant. Of course, for investigators conducting years of longitudinal studies examining numerous questions for which they are not ready to share data, we advise authors to explicitly state in their DAS the reasons for not sharing data. By beginning to capture the true reasons for not sharing data, data availability statements can then begin to be used as support for introducing better protection for and citation of data producers.

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., total DASS-21 score, comprised of 21 individual questions). To explore these different datasets, see the "REDCAP_data.xlsx" and "combined_data.xlsx" files in our Github repository which is archived by Zenodo (DOI:10.5281/zenodo.4011716).

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 association (Odds ratio: 0.91, 95%CI: 0.85–0.96)"). Although 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 example, all study data are shared via Excel files in our Github repository which is archived by Zenodo (DOI:10.5281/zenodo.4011716). This includes the mindfulness trial's quantitative survey data. It also includes qualitative data, which were generated from in-person, semi-structured interviews and coded to examine adherence and problems with using the tools. To demonstrate synthesis-readiness, the relevant responses for qualitative 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. The README file should include a (1) definition of the primary research aims using a relevant alternative framework that addresses the main components of the research which will help others identify and use the research (e.g., PICOTS, TOPICS+M; Haynes, 2006; Johnson & Hennessy, 2019; Stern et al., 2014); (2) a short description of all

the data files included in the project repository; (3) signposts to the detailed data dictionaries for each data file which 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; (5) a description of each data file's relationship to other data files in the repository, and relevant linkage variables, such as participant ID; (6) permanent author contact details; (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.

The README for our mindfulness trial contains the seven main elements necessary for a comprehensive meta-data file; see Figure 2 for a snapshot of these pieces and visit Zenodo to view and download all files: 10.5281/zenodo.4011716.

Archiving and sharing research data

Research data sharing can be done via institutional and non-institutional repositories or alongside published articles (Table 1). We have ordered these sharing options from best practice (most likely to be synthesis-ready) to good practice (still synthesis-ready but may require additional effort to prepare for synthesis).

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, 2020). The main limitation of these repositories is that there are often conditions on how much data can be stored; some charge fees for storage.

For our mindfulness trial example, the repository was created in 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 hosted in the Open Science Framework can have "anonymous" links to remove identifying meta-data as part of double-blind peer-review requirements.

Institutional Repositories.—An institutional repository is one hosted by an academic institution, such as the ICPSR (ICSPR, 2021). An advantage of affiliated repositories is that some provide staff to guide researchers who are new to data sharing. These repositories are less well-known, which may limit the findability of data when compared to other options. The terms 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. This approach reduces data findability 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. If the published article is behind a paywall, then

data attached to it as supplementary files may also be blocked and so data shared as supplemental files may not be fully accessible. 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, workflows, and a full demonstration using an empirical example, in addition to signposting to further literature, to help researchers plan for the resources needed for making research synthesis-ready at project conceptualization. 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 (platform sharing fees) and intangible (time spent preparing data) 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 (Section 5). Finally, some journals have implemented "scoop-protection" policies for research, a system-level solution we hope more journals will implement to address this challenge (Transpose, 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 knowledge about the data, rather than using it to scoop findings 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 meta-data files (Section 5 and Figure 2) should reduce inappropriate use of shared data. Finally, we have provided several different options for maintaining the confidentiality of participants while attending to synthesis-readiness (Section 5).

7. Conclusions

Despite the importance of making research available and synthesis-ready, not all scientific disciplines have embraced data/meta-data sharing and transparency at the same rate (Graf et al., 2020; Hardwicke et al., 2020; Sholler et al., 2019; Vasilevsky et al., 2017). If prevention scientists share research in synthesis-ready forms, the field would see improvements, especially in terms of impact on policy and practice. We presented methods for managing data sharing processes that should make it easier to produce synthesis-ready research.

Changes in research practice that focus on synthesizability will reduce the delay between implementing effective primary study approaches (Nakagawa et al., 2020), and should have a positive impact on translation of prevention science into policy and practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Acabchuk RL, Simon MA, Low S, Brisson JM, & Johnson BT (2021). Measuring meditation progress with a consumer-grade EEG device: Caution from a randomized controlled trial. Mindfulness, 12(1), 68–81. DOI: 10.1007/s12671-020-01497-1

Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jeri K, … & Doré CJ (2013). SPIRIT 2013 statement: Defining standard protocol items for clinical trials. Annals of Internal Medicine; Ann Intern Med. DOI: 10.7326/0003-4819-158-3-201302050-00583

Banks GC, Field JG, Oswald FL, O'Boyle EH, Landis RS, Rupp DE, & Rogelberg SG (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 KE (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 AW, Capers PL, & Kaiser KA (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. DOI:10.1136/bmjopen-2016-012545

Muse Choose. (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 DA, & Rose AK (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 [PubMed: 31851689]

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 [PubMed: 32320428]

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 SR (2016). Gauging the Purported Costs of Public Data Archiving for Long-Term Population Studies. PLOS Biology, 14(4), e1002432. https://doi.org/10/f8kf9w [PubMed: 27058254]

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 LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, & Thompson H (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLOS ONE, 13(5), e0194768. https://doi.org/10/gdhb7j [PubMed: 29719004]

Forstmeier W, Wagenmakers E-J, & Parker TH (2017). Detecting and avoiding likely false-positive findings – a practical guide. Biological Reviews, 92(4), 1941–1968. https://doi.org/10/gdf37n [PubMed: 27879038]

Frankenhuis WE, & Nettle D (2018). Open Science Is Liberating and Can Foster Creativity. Perspectives on Psychological Science,13(4),439–447. DOI:10.1177/1745691618767878 [PubMed: 29961408]

Gaba JF, 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 [PubMed: 32817724]

Gerstner K, Moreno-Mateos D, Gurevitch J, Beckmann M, Kambach S, Jones HP, & 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. 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 EM, Stillman AN, Tingley MW, & Elphick CS (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 JL, Altman D, Bhui K, Booth A, Clark D, Craig P, Eisner M, Fraser MW, Gardner F, Grant S, Hedges L, Hollon S, … 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 [PubMed: 30060763]

Gupta YK, Meenu M, & Mohan P (2015). The Tamiflu fiasco and lessons learnt. Indian Journal of Pharmacology, 47(1), 11–16. https://doi.org/10/gg898k [PubMed: 25821304]

Haddaway NR, & Westgate MJ (2019). Predicting the time needed for environmental systematic reviews and systematic maps. Conservation Biology, 33(2), 434–443. https://doi.org/10/ggnn3p [PubMed: 30285277]

Hardwicke TE, Wallach JD, Kidwell MC, Bendixen T, Crüwell S, & Ioannidis JPA (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 [PubMed: 32257301]

ICPSR. (n.d.). Retrieved February 24, 2021, from https://www.icpsr.umich.edu/web/pages/

Haynes RB (2006). Forming research questions. Journal of Clinical Epidemiology, 59(9), 881–886. 10.1016/j.jclinepi.2006.06.006 [PubMed: 16895808]

Inter-university Consortium for Political and Social Research (2010). Guide to Social Science Data Preparation and Archiving Best Practice Throughout the Data Life Cycle. ICPSR Institute for Social Research University of Michigan.

Johnson BT, & Hennessy EA (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 [PubMed: 31233957]

Kahan BC, & 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 [PubMed: 29720248]

Kerr NL (1998). HARKing: Hypothesizing after the results are known. Personality and social psychology review, 2(3), 196–217. [PubMed: 15647155]

Marshall IJ, Johnson BT, Wang Z, Rajasekaran S, & Wallace BC (2020). Semi-Automated evidence synthesis in health psychology: Current methods and future prospects. Health Psychology Review, 14(1), 145–158. DOI: 10.1080/17437199.2020.1716198 [PubMed: 31941434]

Miyakawa T (2020). No raw data, no science: Another possible source of the reproducibility crisis. Molecular Brain, 13(1), 24. 10.1186/s13041-020-0552-2 [PubMed: 32079532]

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 MW, Gardner F, Grant S, Hedges L, Hollon S, … CONSORT-SPI Group. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 Extension. Trials, 19(1), 407. DOI: 10.1186/s13063-018-2733-1 [PubMed: 30060754]

Nakagawa S, Dunn AG, Lagisz M, Bannach-Brown A, Grames EM, Sánchez-Tójar A, O'Dea RE, Noble DWA, Westgate MJ, Arnold PA, Barrow S, Bethel A, Cooper E, Foo YZ, Geange SR, 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. DOI:10.1038/s41559-020-1153-2 [PubMed: 32203483]

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

Nuijten MB, Hartgerink CHJ, van Assen MALM, Epskamp S, & Wicherts JM (2016). The prevalence of statistical reporting errors in psychology (1985–2013). Behavior Research Methods, 48(4), 1205–1226. https://doi.org/10/f9pdjm [PubMed: 26497820]

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 IV, Borgman CL, & Wofford MF (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. Harvard Data Science Review, 1(2). https://doi.org/10/ggc9n3

Popkin G (2019). Data sharing and how it can benefit your scientific career. Nature, 569(7756), 445–447. https://doi.org/10/gf2v8j [PubMed: 31081499]

Quintana DS (2020). A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. ELife, 9, e53275. https://doi.org/10/gg72dz [PubMed: 32159513]

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

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

Schulz KF, Altman DG, Moher D, & CONSORT Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ (Clinical Research Ed.), 340, c332. 10.1136/bmj.c332

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

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 JP, 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 PD, & 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 AU, 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 [PubMed: 21738610]

Tenopir C, Dalton ED, 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 [PubMed: 26308551]

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

Transpose. (2020). Transpose Database: A database of journal policies on peer review, co-reviewing, and preprinting. https://transpose-publishing.github.io/#/

van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, & Burke DS (2014). A systematic review of barriers to data sharing in public health. BMC Public Health, 14(1), 1144. https://doi.org/10/f6tgv2 [PubMed: 25377061]

Vasilevsky NA, Minnier J, Haendel MA, & Champieux RE (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. 10.1016/j.jbusres.2017.12.043

Walters WH (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

Wilkinson MD, Dumontier M, Aalbersberg Ij. J., Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, 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 [PubMed: 26978244]

Wilson SJ, & Tanner-Smith EE (2014). Meta-analysis in Prevention Science. In Sloboda Z & Petras H (Eds.), Defining Prevention Science (pp. 431–452). Springer US. 10.1007/978-1-4899-7424-2_19

Yenni GM, Christensen EM, Bledsoe EK, Supp SR, Diaz RM, White EP, & Ernest SKM (2019). Developing a modern data workflow for regularly updated data. PLOS Biology, 17(1), e3000125. https://doi.org/10/c82f [PubMed: 30695030]

Zenodo—Research. Shared. (2020). https://about.zenodo.org/
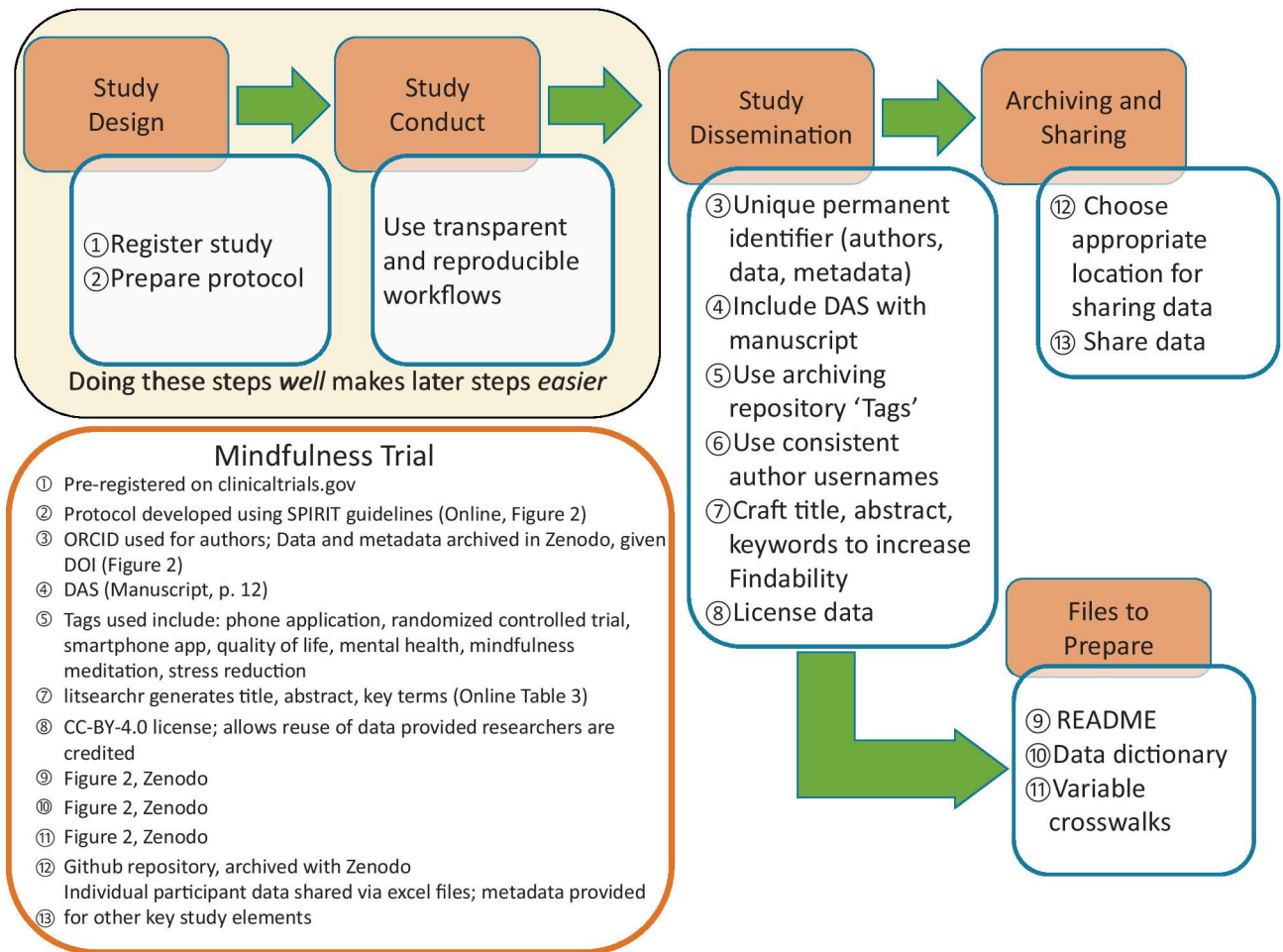
Figure 1.
Best Practices in Generating Synthesis-Ready Research applied to Mindfulness Trial
Note. DAS = Data Availability Statement.

Study description

Taken from the `Study protocol.docx` document, which describes the study approach in detail:

"Young adults in college experience high levels of stress, anxiety, and depression, which can lead to maladaptive behaviors and chronic health issues including inflammation and HPA-axis dysregulation. University mental health centers are seeking effective programs that teach coping skills to self-manage stress, with low stigma, that are easy to implement in large scale. The aim of this protocol is to describe the design of a study aiming to evaluate the effectiveness of a one-month self-guided meditation program for university students comparing various tools to improve mental health, physical health and health behaviors, and explore moderators. Secondary aims include reaching a diverse audience and documenting adverse experiences."

(a)

PICOTS

Population - 140 university undergraduate or postgraduate students studying at an American university.

Intervention - 10 minutes of self-guided meditation daily for four weeks, faciliated by the 10% Happier App, in combination with neurofeedback from the Muse Neurofeedback Tool.

Comparison - 10 minutes of self-guided meditation daily for four weeks, faciliated by the 10% Happier App.

Outcomes - Outcome measures included those related to Personality, Stress, Distress, Resilience (coping), Mindfulness, Emotion regulation, Sleep, Self-Regulation of Eating, Motivation Survey, and Student Adherence (for more details, see Section 2.4 of the `Study protocol.docx` document and the `Instrument List.xlsx` document).

Time - 4 week intervention followed by post-intervention surveys (1-, 6- and 12-months). See `Study Timeline.xlsx` for more details on the tasks performed and the data collected at each timepoint.

Study Design - randomized controlled trial with two intervention groups

(b)

Data source files

All data files are available as Excel spreadsheets in the `data/` subdirectory. The relationships between each data file, and between data files and other information contained in this repository, are described below:

- `RECAP_data.xlsx` (Data dictionary: `Data Dictionary for RECAP Data.pdf`) contains the raw survey data exported from the REDCAP software.
- `External_data.xlsx` (Data dictionary: `Data Dictionary for External Data.xlsx`) contains additional variables on each participant which were not captured by the REDCAP survey. Note some variables, such as C-Reactive protein or cortisol levels, are only available for a subset of participants, based on the availablility of saliva samples for that participant. See Section 2.2.1 in `Study protocol.docx` for more details.
- `Combined_data.xlsx` was created by merging the `RECAP_data.xlsx` and `External_data.xlsx`, and is used as the complete dataset from which composite variables are created using the `create_composite_variables.do` STATA do file. The crosswalk between the individual measures and the composite variables are described in more detail in `Stata Variable Crosswalk.xlsx`
- `Debriefing Interview - Wave 1.xlsx` : contains qualitative reponses from "Wave 1" participants to the debriefing interview questions. *Note: the interview was only done for Wave 1 participants.*

(c)

Permanent contact about this data

Rebeca L. Acabchuk, University of Connecticut, Institute for Collaboration on Health, Intervention, & Policy, Department of Psychological Sciences, United States. ORCID: 0000-0001-8254-2396

Licence

The materials in this repository are licensed under the Creative Commons Attribution 4.0 (CC-BY-4.0) International. The full license file can be found here.

(d)

Figure 2.

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.

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.	[a] 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; Inter-university Consortium for Political and Social Research; UK Data Service	Supported by institution. DOI assigned. Often can get help from staff when uploading data.	Not as well known as other repositories - limits Findability.	[b] Pescosolido et al., 2020, contains scripts and data to replicate all figures/analyses found in the original manuscript.
Supplemental files attached to the online version of the published article	Many file types (.txt, .xlsx, .doc, .pdf) with data or metadata	Likely relatively less work to prepare compared to the other options	No DOI (for supplemental files). Can be closed-access. Potential for data to be lost. Least findable of the other options.	[c] Slep et al., 2020, contains mPLUS code and data [d] Tomczyk et al., 2020, open access, includes data in manuscript and supplemental files

Note. References for the examples listed in the table:

[a] 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

[b] 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

[c] 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

[d] 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