Journal of Clinical Epidemiology 136 (2021) 227–234

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

PRIME-IPD SERIES Part 1. The PRIME-IPD tool promoted verification and standardization of study datasets retrieved for IPD meta-analysis

Omar Dewidar[a,b,*], Alison Riddle[a,b], Elizabeth Ghogomu[a], Alomgir Hossain[b,c], Paul Arora[d], Zulfiqar A Bhutta[e,f], Robert E Black[g], Simon Cousens[h], Michelle F Gaffey[e], Christine Mathew[a], Jessica Trawin[a], Peter Tugwell[i,j,k,l], Vivian Welch[a,b,k,1], George A Wells[b,k,l,1]

[a] *Bruyère Research Institute, University of Ottawa, 85 Primrose Ave, Ottawa, Ontario, K1R 6M1, Canada*
[b] *School of Epidemiology and Public Health, University of Ottawa, 600 Peter Morand Crescent, Ottawa, Ontario, K1G 5Z3, Canada*
[c] *Department of Medicine (Cardiology), The University of Ottawa Heart Institute and University of Ottawa, 40 Ruskin Street, Ottawa, Ontario, K1Y 4W7, Canada*
[d] *Dalla Lana School of Public Health, University of Toronto, 155 College St Room 500, Toronto, Ontario M5T 3M7, Canada*
[e] *Centre for Global Child Health, Hospital for Sick Children, 555 University Ave, Toronto, Ontario, M5G 1X8, Canada*
[f] *Institute for Global Health & Development, Aga Khan University, South-Central Asia, East Africa & United Kingdom, Karachi, Pakistan*
[g] *Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615N Wolfe St Suite E8545, Baltimore, MD, 21205, USA*
[h] *Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine (LSHTM), Keppel Street, London, WC1E 7HT, UK*
[i] *Clinical Epidemiology Program, Ottawa Hospital Research Institute, 501 Smyth Rd, Ottawa, Ontario K1H 8L6, Canada*
[j] *Department of Medicine, University of Ottawa Faculty of Medicine, Roger Guindon Hall, 451 Smyth Rd #2044, Ottawa, Ontario, K1H 8M5, Canada*
[k] *WHO Collaborating Centre for Knowledge Translation and Health Technology Assessment in Health Equity, Bruyère Research Institute, 85 Primrose Ave, Ottawa, Ontario, K1R 6M1, Canada*
[l] *Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, 40 Ruskin St, Ottawa, Ontario, K1Y 4W7, Canada*

Received 30 June 2020; Received in revised form 19 January 2021; Accepted 5 May 2021; Available online 24 May 2021

Abstract

Objectives: We describe a systematic approach to preparing data in the conduct of Individual Participant Data (IPD) analysis.

Study design and setting: A guidance paper proposing methods for preparing individual participant data for meta-analysis from multiple study sources, developed by consultation of relevant guidance and experts in IPD. We present an example of how these steps were applied in checking data for our own IPD meta analysis (IPD-MA).

Results: We propose five steps of Processing, Replication, Imputation, Merging, and Evaluation to prepare individual participant data for meta-analysis (PRIME-IPD). Using our own IPD-MA as an exemplar, we found that this approach identified missing variables and potential inconsistencies in the data, facilitated the standardization of indicators across studies, confirmed that the correct data were received from investigators, and resulted in a single, verified dataset for IPD-MA.

Conclusion: The PRIME-IPD approach can assist researchers to systematically prepare, manage and conduct important quality checks on IPD from multiple studies for meta-analyses. Further testing of this framework in IPD-MA would be useful to refine these steps. © 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Key words: Data replication; Meta-analysis; Individual participant data; Imputation; Research methodology; Data Preparation; Evidence Synthesis; Randomized Trials

Acknowledgments: We would like to thank Celia Holland for being part of our Advisory board. We would also like to thank Bishop Beasley, Nilanthi de Silva, Rebecca Stoltzfus and James Tielsch for contributing data for the network meta-analysis.

Conflict of interest statement: All authors have completed the ICMJE uniform disclosure form and declare: no support from any organization for the submitted work; Ms. Gaffey and Dr. Welch report grants from Bill and Melinda Gates Foundation, during the conduct of the study.

[1] Joint senior authors.

[*] Corresponding author: Tel.: +1-613-501-0632; Fax: +1-613-501-0632.
E-mail address: odewi090@uottawa.ca (O. Dewidar).

https://doi.org/10.1016/j.jclinepi.2021.05.007

What is new?

Key findings

- The multi-step approach that can be used to manage IPD for analysis from multiple studies involves the following stages:
- Processing
- Replication
- Imputation
- Merging
- Evaluation

What this adds to what is known?

- PRIME-IPD provides a formalized step-by-step approach to verify and prepare individual participant data from multiple studies for meta-analysis, thus adding to available guidance on evidence synthesis.

What is the implication and what should change now?

- The synthesis of IPD from multiple trials provides a powerful approach to control for confounding and investigate effect modification at the individual level. However, a principled and systematic way to build the analytic dataset with requisite checks for data quality, is needed to ensure these benefits are realized.
- Further testing of this framework to assess feasibility and applicability to other reviews may refine this model.

1. Introduction

Clinical decision-makers increasingly rely on systematic reviews and meta-analyses because they collate, critically appraise and synthesize all relevant evidence on a particular question [1]. Individual participant data meta-analysis (IPD-MA) is considered the gold standard in systematic reviews since it enables effect modification analyses using individual-level data [2]. IPD-MA is carried out by collecting raw individual participant data from all eligible studies for which the data are available. The data are then pooled and reanalyzed simultaneously [2,3]. IPD-MA has advantages over conventional aggregate data meta-analysis (AD-MA), such as minimizing selective reporting bias and allowing better characterization of subgroups and outcomes as well as data quality assessment [4-6].

IPD are usually acquired by directly contacting the study authors [7]. However, there are multiple barriers to the smooth retrieval of datasets [8,9]. Authors may be hesitant about sharing their datasets due to concerns about how the data will be used, data security and other issues. In the process of including IPD, studies may be subjected to reanalysis. Data sharing hesitancy may stem from apprehensions around having their data scrutinized and re-analyzed [7]. This highlights the importance of developing standardized measures for assessing data quality.

Furthermore, IPD datasets sponsored by industry, such as pharmaceutical and medical device companies, are rarely available and accessible [10]. A systematic review exploring the retrieval of IPD for IPD-MA shows that over 20 years, only 25% of systematic reviews were able to obtain all of the relevant datasets [11]. Over half of the reasons for the unavailability of IPD was the loss of datasets, which highlights the need for improvements in data collection and archiving. Polanin et al. [7] have suggested that using a data-sharing agreement document may alleviate concerns related to data sharing, increasing the likelihood of data sharing and promoting transparent, academic collaboration.

Managing and preparing IPD is resource intensive and time consuming [3,12-14]. IPD datasets differ in their naming conventions, data structure and file formats. Older datasets require even more maintenance as they tend to not be recorded to the current standards. Tudur Smith et al. [15] reported the multiple challenges in data preparation, such as the absence of a summary of variables, data collection in separate files and software incompatibility, resulting in the consumption of extensive amounts of time and resources. Despite the increasing interest in performing IPD-MA and initiatives to improve methods through the Cochrane handbook as well as guidance provided by credible IPD working groups [16-19], there is an absence of a comprehensive and formal approach to collect, verify and analyze the individual level data.

This paper aims to describe the approach we developed and illustrates its value when applied, on an IPD-NMA for mass deworming for children [20,21].

2. Methods

A project advisory group composed of experts in IPD, statisticians, methodologists and systematic reviewers was established to develop a systematic approach to collate and prepare individual participant data for analysis. Prior to developing this approach, we reviewed relevant guidance from the Cochrane Handbook [16], Get Real IPD Working Group, Cochrane Multiple Interventions Group [18] including their library and the Cochrane Methods IPD Meta-Analysis Group [19]. We itemized and categorized components of the relevant guidance and reached consensus on the development of a 5-step approach to prepare the IPD data post-acquisition from study authors. We illustrated the application of this approach to an IPD-NMA on deworming [20,21].

3. Results

Based on the literature and consensus process we developed a five stage systematic approach for the preparation

Table 1. Checklist items for PRIME-IPD tool

PRIME:	Items
Processing	• Convert data into a single format for statistical program of choice • Compare the total number of participants in the acquired datasets to those reported in published studies • Verify the presence of the variables of interest in the acquired dataset • Standardize variable names across datasets • Identify and standardize the measurement scales used to report the variables of interest • Identify and standardize coding for missing values • Identify and correct any implausible values that may result from data conversion
Replication	• Recalculate reported descriptive and summary statistics using the acquired datasets • Calculate the standardized difference to quantitatively assess the difference between the replicated and published results • If the standardized difference is > 10%, investigate and address potential causes
Imputation	• Assess the appropriateness of conducting imputation of missing data using missing data theory • If multiple imputation is conducted, carefully consider the number of imputations to be run
Merging	• Ensure in processing step that variable order and codes are correct • Merge the imputed datasets into a single, pooled dataset, taking into consideration the number of imputed datasets, if appropriate
Evaluation	• Assess continuous variables for normality by residual analysis either visually or by statistical tests • If required, calculate new variables for standardized comparison of effects

and the conduct of an IPD-NMA systematic review. They are:

1. *Processing* of the datasets
2. *Replication* of published data tables
3. *Imputation* of missing data
4. *Merging* datasets
5. *Evaluation* of data heterogeneity

Table 1 provides an overview of the steps undertaken at each stage. The overview is followed by an illustrative example of the methodology using our IPD-NMA of mass deworming interventions for children in low-resource settings [20,21].

3.1. Processing of datasets

The first stage of data processing is to standardize the format of the datasets that will be included in the final IPD analysis. However, several challenges may arise. Acquired datasets may be in different formats (e.g., SAS vs. SPSS), different variable names may be used for the same measure, different scales may be used to report the same measure (e.g., hemoglobin may be reported in grams per liter or grams per decilitre), and some indicators or values may be missing for some individual studies. Missing data may be indicated by different symbols or notations (such as "−99″). Data dictionaries may not be available for all datasets. Consequently, we recommend the following steps in the 'Processing' stage to overcome these challenges:

1. Convert each acquired dataset to a preferred standardized format (e.g., SAS, STATA). The format should be chosen based on facilitating easy data manipulation. This format may or may not be the format used for the eventual analyses.
2. Compare the total number of observations in the received datasets to those reported in the published studies (or global trials registers if publications are not available). In the event of a mismatch, determine the cause of the discrepancies. In event of mismatch, contact the authors to understand reason for discrepancy.
3. Verify that the variables of interest are available in the acquired datasets by referring to accompanying data dictionaries. In their absence, contact the primary authors of the studies for the information required.
4. Create a master list of individual dataset variable names mapped to the variable name of choice. Rename all variables of interest across the datasets to have common variable names.
5. For continuous variables, identify the variables' scales of measurement and identify any datasets that may need to have values converted to the preferred standard using appropriate conversion formula(e). Determine whether the categories of the categorical variables need to be regrouped or separated into dummy variables.
6. Identify any missing values in the datasets and how they are identified in the dataset (e.g., blank cells, symbols). Confirm that the blank cells are missing values and not due to a conversion error by comparing the percentage of missing values per variable in the acquired and converted datasets and standardize across datasets. Similar considerations may exist for data considered not applicable.

3.2. Replication of published data tables

The second step is to replicate the data tables reported in the published studies. Since reproducibility is an anchor in scientific research [22-24], it is essential to check that the processed datasets are consistent with the analyzed datasets in the published papers. This step will provide an additional check on data quality and fidelity to the acquired study and increase confidence that the datasets were processed correctly. Discrepancies between the replicated and published results are often expected [25,26]. Challenges in the replication process include discrepancies in the number of participants or units of analysis between the published paper and the acquired datasets and lack of reporting on statistical methods. The following steps are proposed to assist in minimizing these challenges:

1. Calculate and compare the descriptive statistics from the processed datasets to the published results. For example, the percentage of females enrolled, age of participants, and pre-existing health conditions.
2. Calculate and compare baseline and endline summary statistics for the outcomes of interest from the processed datasets to the published results using the same analytic methods reported in the published article.
3. Calculate the standardized difference between the descriptive and summary statistics of the published studies and the replicated results. We referred to the absolute standardized difference criterion of 10% proposed to assess baseline imbalance to assess the magnitude of difference between replicated and published results. We chose the criterion of 10% as an indicator of discrepancy between published and replicated results based on previously proposed thresholds [27-30]. The standardized difference can be calculated as follows:

$$\frac{(difference\ between\ replicated\ and\ published\ results)}{\sqrt{Variance}\ (difference\ between\ replicated\ and\ published\ results)}$$

Note: Variance is calculated assuming independence of replicated and published results.

3.3. Imputation of missing data

Missing data are inevitable in clinical research [31]. Complete case analysis, ignoring participants with missing data, has the potential to bias results [32] and can reduce a study's precision and power due to a smaller sample size. Imputation may be considered to redress missing data, depending on the amount and type of missingness in the processed datasets and according to missing data theory [33]. If multiple imputation is implemented, carefully consider the number of imputations to be run [34], taking into consideration that a greater number of imputations will result in longer computing time [32].

3.4. Merging datasets

The merging of datasets in this context refers to the vertical merging of rows (or observations) of two or more datasets. All datasets to be combined in the merge step should already have the same variables (or columns) following the processing step. Different statistical programs will have different names for this command, or multiple ways that datasets may be merged. For example, in SAS you may use concatenation in the DATA step command, or the APPEND procedure (SAS Institute Inc., Cary, NC, USA). Readers should follow the guidance for merging provided by their statistical software program, as specific steps can vary. It is important to ensure that you can identify from which original study each observation belongs after the merge step. This can be done by creating a variable for study name. Alternatively, in Stata, you may employ the "generate" option [35] to create a variable identifying from which dataset each observation originally came. Observations from imputed datasets will also need to be correctly labelled according to their original study and imputation number.

3.5. Evaluation of data heterogeneity

Prior to the conduct of a pooled analysis, an assessment of the merged dataset's heterogeneity and distribution may be explored to inform statistical methods and interpretation of results. Further, authors may need to calculate new variables for the standardized comparison of effects. We suggest the following:

1. Test data distribution by residual analysis for continuous variables either visually, by preparing bar charts, or by parametric statistical tests [36]. Comparisons can be implemented between study arms to appraise the randomization of participants in each group and identify differences between study groups.
2. Create new variables needed for analysis (e.g., "dummy variables" for categorical variables). This step is needed if there are any variables which need to be calculated based on existing variables in the merged dataset (e.g., body mass index may be calculated using existing data on the height, weight, age and sex of participants).

4. PRIME application

We report our experience using PRIME-IPD for preparing data for an individual participant data network meta-analysis of mass deworming interventions for children in low-resource settings [20,21] as an exemplar in Table 2. The appended table shows the value-added of each step in verifying and standardizing the acquired data for use in an IPD-NMA.

5. Discussion

This paper details a methodology for the preparation of data for IPD-MA composed of five steps: Processing, Replication, Imputation, Merging and Evaluation. Standardization of included datasets is performed in the processing step, followed by verification of datasets through data replication. To deal with missing data, we propose imputation if appropriate, according to missing data theory. Following the merging of the processed datasets and prior to conducting analyses using the merged dataset, we suggest assessing heterogeneity across the variables in the evaluation step and creating any new variables that are required for analysis. Many aspects within PRIME-IPD will help formulate the Statistical Analysis Plan (SAP) for IPD.

Table 2. Application of PRIME-IPD in the context of a deworming systematic review

PRIME:	Problem	Application
Processing	Incomplete and missing data dictionaries	We used a list of analysis variables to request data since it identified which variables we needed and reviewed the dataset files along with the data dictionaries. Correspondence with authors was helpful in preparing datasets lacking dictionaries.
	Identification of missing variables of interest	We documented the choice of outcome measures for studies that collected data at multiple time points and identified four out of 11 studies which did not report the primary outcomes of interest in their published manuscripts, but they did collect this data and provided it in their dataset.
	Use of different measurement methods	We evaluated the measurements for helminth egg counts were provided. We identified various measurement methods used between authors in terms of the number of egg samples taken and how they were collected. We selected the most common method and standardized it across all included studies.
	Identification of conversion errors	We identified the presence of implausible values that required conversion before analysis such as zeros coded 0.99 and 9999.
Replication	Inexact number of participants in the datasets compared to reported	The authors provided full datasets, including children who were excluded from the analysis due to missing baseline measures (e.g., missing stool samples). Replication allowed us to verify that these children were excluded from the analyses in the published papers.
	Incorrect treatment labels	By means of replication, we found that the labels in the dataset from authors did not match the labels in the published paper. Correspondence with the authors allowed us to correct these labels and replicate the analyses
	Uncorrected variables in the provided datasets	Hemoglobin concentration need to be corrected if measured in individuals living in areas 1000 m above sea level, since lower oxygen levels, result in higher hemoglobin concentrations in the blood. Hemoglobin was not adjusted for in two studies' datasets which were carried out in areas 1000 m above sea level, so the Hemoglobin concentration values obtained when replicating were larger than the reported [37].
Imputation	Studies with missing data	For each study included in the IPD analysis, we calculated the percentage of missing data for each variable of interest. Consequently, we assessed the distribution of the missing variables to assess if imputation was appropriate. We imputed the eligible studies that had less than 50% of missing data and assumed data were missing at random, creating five imputed datasets per study. We used complete case analysis for studies with more than 50% of missing data as part of sensitivity analyses only.
Merging	Correctly combining multiple datasets	A separate variable was created to identify each observation's original study and imputation number (ranging from one to five). We sorted datasets by that identifier and used MERGE used the command in SAS (9.4) to combine the imputed datasets into a new dataset.
Evaluation	New variable calculation	Growth standards have varied over the years. We used WHO anthropometric software to calculate BMI for age, weight for age and other growth standards in relatively older studies to combine with the other studies. The Anthropometric calculator in the software also operates similar SAS by tagging implausible weight and height values.

Subsequent to data preparation, synthesizing study data is needed to assess the intervention effects. This step bears its own series of barriers and challenges with guidance provided elsewhere [6,13,17,38].

The five step approach of PRIME-IPD is a comprehensive composite of previous research methods and guidance for IPD. The Cochrane Handbook version 5 [16] highlights the importance of recoding variables during data preparation but does not detail procedures to prepare the dataset for IPD analysis. The "get-real" review conducted by Debray et al. provides insight on how to distinguish between different missing data scenarios but does not provide suggestions when considered, may improve the robustness of the imputation process [17].

An important aspect of our approach is in including a data replication step, which can help verify what the authors report in their published studies and identify any errors present in the processed datasets. Replicating the acquired studies' descriptive and summary statistics helps to identify critical assumptions made by the original investigators and data inconsistencies in the acquired datasets. This process adds to the robustness of the IPD analysis conducted by the investigators. There are additional proposed benefits to re-analyzing the datasets as conducted by the original investigators such as ensuring complete, accurate and unbiased reporting of results [39]. However, the additional time and cost that may be incurred to conduct the replication should be considered [40]. The replication process can become unwieldy with a large number of studies. We also acknowledge that the PRIME-IPD methodology is a lengthy process. However, based on firsthand experience, we found the benefits outweigh the costs because we were able to identify and correct data problems before pooling and data synthesis.

The success of conducting IPD-MA does not solely depend on the preparation of datasets for analysis, but heavily depends on the ability to retrieve datasets from authors. There are several challenges to accessing IPD, as it is often unavailable even upon request from authors [41-43]. Less than 50% of IPD-MA systematic reviews published between 1987 and 2015 succeeded in retrieving at least 80% of their selected studies [7]. Therefore, there is a crucial need to build confidence and trust among investigators using data sharing agreements, which have been shown to increase the likelihood of a response [7,44] and using investigator collaboratives [7]. Improving IPD access coincides with initiatives to make data available through online data repositories (public or private) such as Vivli [45] and OpenTrials [46, 47-51]. The role of data repositories in facilitating IPD analysis remains limited in terms of the curation of datasets for analysis. Investigators are requested to upload dictionaries, but they are usually incomplete, and datasets lack organization with major heterogeneity between studies [52]. The PRIME-IPD approach overcomes these hurdles when dealing with several datasets through providing a systematic approach to preparing data for analysis, including verification of terms with authors if needed. Data-sharing repository services may address this limitation in the future by unifying policies and systems.

6. Conclusion

PRIME-IPD proposes a systematic approach to the preparation and verification of individual participant datasets. Combining PRIME-IPD with best practices in acquiring datasets from authors, such as the use of data-sharing agreements, and offering appropriate acknowledgement and incentives, may improve efficiency in conducting IPD analysis. Nonetheless, the PRIME-IPD approach requires further testing in different settings and may be require adaptations in specific scenarios.

Acknowledgments

Patient and public involvement: No patients or public were involved in the conduction or the design of this project.

Data sharing: There is no data to be shared.

Ethics approval: The systematic review was approved by the Bruyère Research Institute and SickKids research ethics boards.

Contribution: All authors contributed to the development of the methodology. AR, OD, EG, CM, AH, JT, VW and GW applied and refined the methodology. OD and AR took the lead in writing the manuscript with oversight from VW and GW. All authors provided critical feedback and helped shape the research, analysis and manuscript. VW and GW are joint senior authors.

Transparency declaration: The manuscript's guarantor (OD) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Author statement

Omar Dewidar: Conceptualization, Methodology, Roles/Writing - original draft; Writing - review & editing **Alison Riddle:** Conceptualization, Methodology, Roles/Writing - original draft; Writing - review & editing, **Elizabeth Ghogomu:** Conceptualization, Methodology, Writing - review & editing **Alomgir Hossain:** Conceptualization, Methodology, Writing - review & editing **Paul Arora**: Conceptualization, Methodology**,** Writing - review & editing **Zulfiqar A Bhutta:** Conceptualization, Methodology, Writing - review & editing **Robert E Black:** Conceptualization, Methodology, Writing - review & editing **Simon Cousens:** Conceptualization, Methodology, Writing - review & editing **Michelle F Gaffey:** Conceptualization, Methodology, Writing - review & editing **Christine Mathew:** Data curation **Jessica Trawin:** Data curation **Peter Tugwell:** Conceptualization, Methodology, Writing - review & editing **Vivian Welch:** Conceptualization, Writing - review & editing, Supervision **George A Wells:** Conceptualization, Writing - review & editing, Supervision

Funding

This research was funded by the Bill and Melinda Gates Foundation [OPP1140742].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.05.007.

References

[1] Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011;128:305–10 PubMed PMID: 21701348; PubMed Central PMCID: PMCPMC3124652. doi:10.1097/PRS.0b013e318219c171.

[2] Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. JAMA 2015;313:1657–65 Epub 2015/04/29PubMed PMID: 25919529. doi:10.1001/jama.2015.3656.

[3] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. Bmj 2010;340:c221 Epub 2010/02/09PubMed PMID: 20139215. doi:10.1136/bmj.c221.

[4] Levis B, Benedetti A, Levis AW, Ioannidis JPA, Shrier I, Cuijpers P, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the patient health questionnaire-9 depression screening tool. Am J Epidemiol 2017;185:954–64 PubMed

PMID: 28419203; PubMed Central PMCID: PMCPMC5430941. doi:10.1093/aje/kww191.

[5] Vale CL, Rydzewska LH, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. Bmj 2015;350 h1088. Epub 2015/03/10PubMed PMID: 25747860; PubMed Central PMCID: PMCPMC4353308. doi:10.1136/bmj.h1088.

[6] Stewart LA, Tierney JF. To IPD or not to IPD?:Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76–97 PubMed PMID: 11868447. doi:10.1177/0163278702025001006.

[7] Polanin JR, Williams RT. Overcoming obstacles in obtaining individual participant data for meta-analysis. Res Synth Methods 2016;7:333–41 Epub 2016/05/28PubMed PMID: 27228953. doi:10.1002/jrsm.1208.

[8] Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. Psychol Methods 2009;14(2):165–76 PubMed PMID: 19485627. doi:10.1037/a0015565.

[9] Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE 2013;8(7):e67332 Epub 2013/07/23PubMed PMID: 23935830; PubMed Central PMCID: PMCPMC3720779. doi:10.1371/journal.pone.0067332.

[10] Murugiah K, Ritchie JD, Desai NR, Ross JS, Krumholz HM. Availability of clinical trial data from industry-sponsored cardiovascular trials. J Am Heart Assoc 2016;5:e003307 Epub 2016/04/20PubMed PMID: 27098969; PubMed Central PMCID: PMCPMC4859296. doi:10.1161/JAHA.116.003307.

[11] Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. Bmj 2017;357 j1390. Epub 2017/04/07PubMed PMID: 28381561; PubMed Central PMCID: PMCPMC5733815. doi:10.1136/bmj.j1390.

[12] Clarke MJ. Individual patient data meta-analyses. Best Pract Res Clin Obstet Gynaecol 2005;19:47–55 Epub 2004/12/13PubMed PMID: 15749065. doi:10.1016/j.bpobgyn.2004.10.011.

[13] Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane working group. Stat Med 1995;14(19):2057–79 PubMed PMID: 8552887. doi:10.1002/sim.4780141902.

[14] Abo-Zaid G, Sauerbrei W, Riley RD. Individual participant data meta-analysis of prognostic factor studies: state of the art? BMC Med Res Methodol 2012;12:56 Epub 2012/04/24PubMed PMID: 22530717; PubMed Central PMCID: PMCPMC3413577. doi:10.1186/1471-2288-12-56.

[15] Tudur Smith C, Nevitt S, Appelbe D, Appleton R, Dixon P, Harrison J, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. Trials 2017;18:319 Epub 2017/07/17PubMed PMID: 28712359; PubMed Central PMCID: PMCPMC5512949. doi:10.1186/s13063-017-2067-4.

[16] LA S, JF T, Cochrane MC. Handbook for systematic reviews of interventions, Version 5.1.0. The Cochrane Collaboration; 2011. ed.

[17] Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RH, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods 2015;6:293–309 Epub 2015/08/20PubMed PMID: 26287812; PubMed Central PMCID: PMCPMC5042043. doi:10.1002/jrsm.1160.

[18] Cochrane Methods Comparing Multiple Interventions: The Cochrane Collaboration. 2021 Available from: July 16, 2020, https://methods.cochrane.org/cmi/.

[19] Cochrane Methods IPD Meta-analysis Group: The Cochrane Collaboration. 2021 Available from: July 16, 2020, https://methods.cochrane.org/ipdma/.

[20] Welch VA, Hossain A, Ghogomu E, Riddle A, Cousens S, Gaffey M, et al. Deworming children for soil-transmitted helminths in low and middle-income countries: systematic review and individual participant data network meta-analysis. J Development Effectiveness 2019;11:288–306. doi:10.1080/19439342.2019.1691627.

[21] Welch VA, Ghogomu E, Hossain A, Riddle A, Gaffey M, Arora P, et al. Mass deworming for improving health and cognition of children in endemic helminth areas: a systematic review and individual participant data network meta-analysis. Campbell Systematic Reviews 2019;15:e1058. doi:10.1002/cl2.1058.

[22] McNutt M. Reproducibility. Science 2014;343:229 PubMed PMID: 24436391. doi:10.1126/science.1250475.

[23] Makel MC, Plucker JA, Hegarty B. Replications in psychology research: how often do they really occur? Perspect Psychol Sci 2012;7:537–42 PubMed PMID: 26168110. doi:10.1177/1745691612460688.

[24] Simons DJ. The Value of Direct Replication. Perspect Psychol Sci 2014;9:76–80 PubMed PMID: 26173243. doi:10.1177/1745691613514755.

[25] Klein RA, Ratliff KA, Vianello M, Adams RB, Bahník Š, Bernstein MJ, et al. Investigating variation in replicability. Soc Psychol 2014;45:142–52. doi:10.1027/1864-9335/a000178.

[26] Nosek BA, Errington TM. Making sense of replications. Elife 2017;6 Epub 2017/01/19PubMed PMID: 28100398; PubMed Central PMCID: PMCPMC5245957. doi:10.7554/eLife.23383.

[27] Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics - Simulation and Computation 2009;38:1228–34. doi:10.1080/03610910902859574.

[28] Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg 2007;134:1128–35 PubMed PMID: 17976439. doi:10.1016/j.jtcvs.2007.07.021.

[29] Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Stat Med 2008;27:2037–49 PubMed PMID: 18038446. doi:10.1002/sim.3150.

[30] Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54:387–98 PubMed PMID: 11297888.

[31] Dong Y, Peng CY. Principled missing data methods for researchers. Springerplus 2013;2:222 Epub 2013/05/14PubMed PMID: 23853744; PubMed Central PMCID: PMCPMC3701793. doi:10.1186/2193-1801-2-222.

[32] Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393 Epub 2009/06/29PubMed PMID: 19564179; PubMed Central PMCID: PMCPMC2714692. doi:10.1136/bmj.b2393.

[33] Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10:585–98 PubMed PMID: 2057657.

[34] Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. Multivariate Behav Res 1998;33:545–71 PubMed PMID: 26753828. doi:10.1207/s15327906mbr3304_5.

[35] StataCorp. Stata 16 base reference manual. College Station, TX: Stata Press; 2019.

[36] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10:486–9 Epub 2012/04/20PubMed PMID: 23843808; PubMed Central PMCID: PMCPMC3693611. doi:10.5812/ijem.3505.

[37] Dirren H, Logman MH, Barclay DV, Freire WB. Altitude correction for hemoglobin. Eur J Clin Nutr 1994;48:625–32 PubMed PMID: 8001519.

[38] Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: guidance on Their Use. PLoS Med 2015;12:e1001855 Epub 2015/07/22PubMed PMID: 26196287; PubMed Central PMCID: PMCPMC4510878. doi:10.1371/journal.pmed.1001855.

[39] Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ, et al. Reanalyses of randomized clinical trial data. JAMA 2014;312(10):1024–32 PubMed PMID: 25203082. doi:10.1001/jama.2014.9646.

[40] Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in. BMJ 2018;360 k400. Epub 2018/02/13PubMed PMID: 29440066; PubMed Central PMCID: PMCPMC5809812. doi:10.1136/bmj.k400.

[41] Cohen B, Vawdrey DK, Liu J, Caplan D, Furuya EY, Mis FW, et al. Challenges Associated With Using Large Data Sets for Quality Assessment and Research in Clinical Settings. Policy Polit Nurs Pract 2015;16:117–24 Epub 2015/09/08PubMed PMID: 26351216; PubMed Central PMCID: PMCPMC4679583. doi:10.1177/1527154415603358.

[42] Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract 2017;36:3–11 Epub 2017/03/31PubMed PMID: 28392994; PubMed Central PMCID: PMCPMC5331970. doi:10.23876/j.krcp.2017.36.1.3.

[43] Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353 i3140. Epub 2016/06/22PubMed PMID: 27334381; PubMed Central PMCID: PMCPMC4916924. doi:10.1136/bmj.i3140.

[44] Wolfe N, Gøtzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. Syst Rev 2013;2(1):31. doi:10.1186/2046-4053-2-31.

[45] Vivli [cited 2020 02/06]. 2021 Available from: July 16, 2020, https://vivli.org/about/overview-2/.

[46] OpenTrials [cited 2020 02/06]. 2021 Available from: July 16, 2020, https://opentrials.net/.

[47] Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Trials 2010;11:9. doi:10.1186/1745-6215-11-9.

[48] Institute of Medicine Sharing clinical trial data, maximizing benefits, minimizing risk. Washington, DC: National Academies Press (US); 2015.

[49] Vickers AJ. Sharing raw data from clinical trials: what progress since we first asked "Whose data set is it anyway?". Trials 2016;17:227 Epub 2016/05/04PubMed PMID: 27142986; PubMed Central PMCID: PMCPMC4855346. doi:10.1186/s13063-016-1369-2.

[50] Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. N Engl J Med 2013;369:1651–8 Epub 2013/10/21PubMed PMID: 24144394. doi:10.1056/NEJMhle1309073.

[51] Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open 2017;7:e018647 Epub 2017/12/14PubMed PMID: 29247106; PubMed Central PMCID: PMCPMC5736032. doi:10.1136/bmjopen-2017-018647.

[52] Banzi R, Canham S, Kuchinke W, Krleza-Jeric K, Demotes-Mainard J, Ohmann C. Evaluation of repositories for sharing individual-participant data from clinical studies. Trials 2019;20:169 Epub 2019/03/15PubMed PMID: 30876434; PubMed Central PMCID: PMCPMC6420770. doi:10.1186/s13063-019-3253-3.