Journal of Physiotherapy 67 (2021) 224–227

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Appraisal

Research Note: Individual participant data (IPD) meta-analysis

A systematic review is a robust method with which to search for, identify, extract, and synthesise evidence from individual studies to answer a specific research question.[1] Meta-analysis is a statistical analysis approach used in some systematic reviews to combine quantitative information across studies, in order to produce overall summaries of the evidence (eg, of a treatment's effect). Meta-analyses are most often conducted using data that has been extracted from peer-reviewed publications included in systematic reviews; such data are often called aggregate data, since they represent information combined across all participants in a particular study. The extracted data typically include a small number of data pieces from each study, such as the change in pain (mean, standard deviation) between treated and untreated study groups, which would allow a treatment effect estimate and its confidence interval to be calculated. An aggregate data meta-analysis is a useful approach with which to summarise the average overall effect of a treatment. However, having only aggregated group data limits the analyses that are possible, and in particular makes it problematic to examine relationships where individual participant-level covariates are of interest. To address this, another option to synthesise evidence across studies is to use the original, participant-level study data, using an approach called individual participant data (IPD) meta-analysis.

This Research Note describes the steps involved in an IPD meta-analysis, explains when this research approach is most useful, and discusses key advantages, challenges and potential future directions. Table 1 provides definitions of some key terms. Although this Research Note focuses on meta-analysis of randomised trials evaluating treatment effectiveness, most points apply more broadly.

What is an IPD meta-analysis project and what are the steps?

IPD meta-analysis projects are often considered to be the 'gold standard' for synthesising evidence across studies.[2] Table 2 highlights key differences in the steps of a systematic review with IPD meta-analysis and a systematic review with a traditional aggregate data meta-analysis. With an IPD project, raw individual-level data about each participant (including baseline demographics, health conditions, prognostic factors and other relevant characteristics, as well as outcome data) are obtained for each published study, then cleaned, harmonised and synthesised together. The steps involved in conducting a high-quality IPD project are described below, and additional detail is provided in a forthcoming handbook.[3]

Search and selection

A well conducted IPD meta-analysis is part of a systematic review with clear research questions, inclusion criteria, comprehensive search, systematic selection, transparent study-level data extraction and risk of bias assessment. Selection criteria for systematic reviews using IPD are similar to traditional reviews, typically including population, intervention, comparison and outcome (PICO) criteria, and may include additional criteria such as sample size or risk of bias requirements (based on published information). Any reasons for study exclusion should be clearly documented.[4]

Accessing individual participant data

IPD can be collected through direct contact with primary study authors with the use of data-sharing agreements or through data-sharing platforms (eg, disease-specific data repositories, journal websites). It is preferable to deal directly with study authors, as the IPD in a data-sharing platform may be incomplete (eg, not all outcomes provided) and unavailable to analyse and store locally. IPD projects should clearly describe how data were requested, collected and managed (eg, list and define all study-level and participant-level data that were sought, including baseline and follow-up information).

Data verification, manipulation and replication

Data from each primary study are evaluated, once received, to confirm comparability to the study publication(s) for descriptive baseline data, range of included variables and missing observations. Any discrepancies or missing information from those presented in the original publication(s) should be clarified with the primary study authors. Risk of bias classifications should also be updated at this stage, based on the IPD itself.

Study variable mapping

Once IPD are separately finalised for each study, they need to be harmonised as far as possible, so that included variables are consistently defined across studies. The homogeneity of the resulting master dataset depends on data availability of common measurements in the primary studies. Whenever possible, variables measured continuously should be kept in their continuous data form. Any standardisation and translation process should be described in detail in a data management plan.

Analysis

IPD meta-analyses are planned and described a priori and include descriptive, analytic and, possibly, exploratory analyses. There are two main approaches to conduct IPD meta-analysis: one-stage or two-stage. In a two-stage IPD meta-analysis, an estimate of the effect of interest (eg, treatment effect or interaction between treatment and participant-level characteristic) is first separately calculated for each trial; these can be presented in forest plots and then (in the second stage) combined in a similar manner to a traditional meta-analysis to produce an overall summary. In one-stage IPD meta-analysis, a single multi-level model is estimated based on the IPD for all studies together, while accounting for the clustering of data within each trial. The one-stage approach is appealing as it simultaneously estimates multiple parameters in a single step and allows more flexibility in modelling assumptions. The two-stage approach is appealing as it

https://doi.org/10.1016/j.jphys.2021.04.001
1836-9553/© 2021 Australian Physiotherapy Association. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
Definitions of key terms related to individual participant data meta-analysis.

Term	Definition
Meta-analysis	Statistical technique for combining quantitative evidence across multiple studies
Aggregate data	Information averaged or summarised across a group of participants in a study, such as the overall treatment effect, mean age or the proportion of females Typically as reported in publications of the primary studies Traditionally extracted and used for meta-analysis
Individual participant data (IPD)	Raw participant-level data collected from the primary studies Made available for IPD meta-analysis, most commonly from the study researchers
One-stage, two-stage	Different statistical approaches to conducting the meta-analysis of IPD Two-stage: obtain estimates (aggregate data) for each study separately, then pool these across studies One-stage: single, usually hierarchical or multi-level, statistical model synthesises data from all IPD in a single step, whilst accounting for clustering of participants within trials
Treatment effect modification	When the treatment effectiveness size differs based on another factor (eg, participant characteristic)

Table 2
Key differences between a traditional aggregate data meta-analysis and an individual participant data (IPD) meta-analysis for treatment effectiveness questions.

Characteristic	Traditional aggregate data meta-analysis	Differences when using an IPD meta-analysis
Study identification	Systematic review with comprehensive search to identify all studies that answer the research question	None – usually IPD meta-analysis projects also use a systematic review with comprehensive search to identify all studies that answer the research question.
Study inclusion	All available studies, including published and (where identifiable) unpublished studies	In addition, due to the painstaking nature of retrieving IPD, the project may apply additional selection criteria to identify a subset of the studies (those low risk of bias, larger sample size) to prioritise for IPD collection.
Data collected and analysed	Aggregate data extracted from each study published report, or sometimes requested from study authors directly	Original participant data are requested, obtained and cleaned from each study in close collaboration with the original study researchers.
Analysis aims	To summarise the average treatment effect across all studies, for each outcome reported To quantify any between-study heterogeneity in the treatment effect, and potentially assess subgroups defined by study or population-level characteristics (eg, mean age)	In addition, IPD meta-analyses usually aim to model patient-level relationships; in particular, to obtain treatment effects adjusted for prognostic factors, and to examine potential treatment effect modification by participant-level characteristics (treatment-covariate interactions). Furthermore, any outcomes or time points available in the IPD can be considered, regardless of whether they were reported upon publication.

enables familiar meta-analysis methods and the presentation of individual study results, and so eases interpretation. When they use the same estimation methods and assumptions, one-stage and two-stage approaches garner generally similar results.[5] When most studies are small (ie, patients and events are few), a one-stage approach is more exact and therefore preferable. When examining potential participant-level modifiers of treatment effect, this is performed by estimating treatment-covariate interactions that quantify how treatment effectiveness varies as participant-level characteristics change. Riley et al describe how to perform this in a one-stage or two-stage approach,[6] and stress the importance of separating within-trial from across-trial information to avoid aggregation bias.

Reporting

The PRISMA reporting guideline and checklist extension for IPD (PRISMA-IPD) should be followed for reporting.[7]

When is an IPD meta-analysis the best approach to address a health research question?

IPD studies are time and resource intensive; it often takes more than 2 years to obtain, clean and synthesise IPD. The best value for this considerable expense is when: the aggregate data needed to answer the research question are not reported/available in the primary study publications, and/or the IPD are needed to go beyond the analyses or aims of the original primary studies. These situations particularly occur in health research questions related to diagnosis, prognosis or treatment effectiveness, as presented in Table 3, where participant-level covariates and relationships are of interest for modelling.

Primary study data may be missing for traditional meta-analysis if a study was not published, or results (or outcomes) were partially or selectively published. In the latter situation, a larger primary study sample (eg, if only a study population subset was published) or additional outcome measures, measured but not reported, may be available by collecting and analysing IPD.

Larger amounts of more homogeneous data may be available for syntheses using IPD if exposure, covariate or outcome measures selected for presentation in publications differ across studies. If study results are reported across relevant primary studies using different scales, follow-up time points or cut-off values, these cannot easily be synthesised using traditional meta-analysis and use of IPD can potentially allow for standardisation of these measures (eg, Levis 2020; Holden 2021).[8,9]

Additionally, analyses that require large sample sizes, such as development and testing of prediction models (eg, Hudda 2019[10]) or analyses of treatment effect modifiers (eg, Hayden 2019[11]), can benefit from the use of IPD to increase power. These types of analyses investigate individual characteristics at a participant level in ways beyond what is typically planned and feasible for the primary studies. Very few randomised controlled trials are designed to detect treatment effect modifiers.[12] An adequately powered primary study (ie, trial) to assess treatment effect modification or compare multiple treatment approaches would need to be extremely large (and more expensive than IPD).[13,14]

What are the advantages of an IPD meta-analysis project?

There are several advantages of systematic reviews using IPD over traditional meta-analyses based on aggregate data.[15] Systematic reviews including IPD may improve data availability and quality to reduce publication bias. Re-analyses of the original raw participant

Table 3
Examples of individual participant data (IPD) meta-analysis projects for different types of questions.

Question type	Research question and use of IPD
Diagnostic test accuracy	Comparison of two depression scales: assessment of diagnostic accuracy in a manner that was not analysed in the primary studies (heterogeneous tool measurements; multiple cut-offs rather than only what was published; no previous meta-analysis and only two primary studies were available to answer this question).[8]
Predictive modelling	Development and validation of a prediction model for fat mass in children: IPD from four primary studies provided sufficient data to develop and validate a model using routinely available risk factors in existing datasets. The final model including height, weight, age, sex and ethnicity was reported to have high predictive accuracy.[10]
Overall prognosis and prognostic factors	Prognostic factors for non-traumatic adolescent knee pain: IPD enabled selection and harmonising of prognostic factors and both short-term and long-term outcomes across studies that had not been reported in a homogeneous way.[9]
Intervention effectiveness	Exercise treatment for chronic LBP: this IPD meta-analysis was able to identify potential treatment effect modifiers for exercise therapy that were not available in the primary studies. There was consistent evidence that heavy work demands and use of pain medications each modify the effectiveness of exercise therapy compared with other interventions.[11]

data can allow more consistent analyses or data presentation, participant subgroups, or definition of outcomes, and can reduce risk of bias concerns. Also, unpublished trials or unreported/incomplete data (and outcomes) can be included if IPD are available for these studies.

IPD meta-analysis can enable better investigation of subgroups and treatment effect modifiers, considering effects in participants with different characteristics (disentanglement of participant-level and study-level sources of heterogeneity in treatment effect). The greater power for these types of analyses is difficult to achieve from individual studies or from aggregate data. Traditional meta-analysis can explore treatment-covariate interactions by meta-regression on study-level characteristics. These analyses are limited to study/population-level data (with limited variation at the aggregate level – eg, mean age) and are prone to study-level confounding (causing aggregation bias).[6] IPD gives more power, a larger number of individual-level covariates and a wider range of covariate values. IPD meta-analysis enables direct derivation of desired information at different time points, the correlation amongst multiple outcomes or time-points to be accounted for in the analysis (which can lead to efficiency gains),[16] the modelling of continuous outcomes and variables on their continuous scale (thereby avoiding the use of arbitrary cut-off values), and the adjustment for prognostic factors and effect modifiers to improve consistency for network meta-analysis.[17] All of these advantages have the potential to provide better data, which can enhance clinical decision-making.

The collaborative nature of IPD projects provides the advantage of a larger team, with input and engagement on the project from experienced trialists. Furthermore, this may enable more unpublished data to be included, facilitate prospective planning of multisite and multi-country trials, support standardisation of prognostic factor measurement, and foster sharing of data through accessible repositories.[18]

What are the challenges of an IPD meta-analysis project?

A potential limitation of IPD studies is the inconsistent availability and measurement of some individual variables. This can limit the ability to include prognostic factors or to assess all potential treatment effect modifiers with the most valid and reliable measures.

IPD meta-analysis projects may not be able to retrieve all study data; selection bias/availability bias may be a challenge if relevant data are not available for all trials.[19] To address this challenge, where possible, IPD meta-analysis should compare or combine data from the included IPD trials with aggregate data from other trials.[20] IPD meta-analysis should report the proportion of eligible included studies and reasons for unavailability. The reader should consider if bias was likely to be introduced by unavailable studies and/or data.

An additional challenge of the IPD approach, acknowledged by all researchers with experience in this study design, is the considerable amount of time and effort that is involved in gathering, testing and compiling data from individual studies.[21] Relative to the often-insurmountable cost to prospectively collect adequate data for effect modification analyses, IPD analyses are very efficient.

The future of IPD meta-analysis

Most systematic reviews summarise aggregate data.[22] However, availability of IPD is likely to increase in the future due to an increase in data repositories, data-sharing initiatives and expectations from research funders. The *International Committee of Medical Journal Editors* (ICMJE) member journals required a data-sharing statement as of July 2018, and a data-sharing plan for trials registered after January 2019. While they do not require data sharing, they have established minimal requirements intended to move the medical research field toward the goal of universal data sharing. In addition to top-down mandates, there is increased recognition of the academic value of dataset development and sharing. Data sharing may increase the feasibility and conduct of IPD meta-analysis through more readily available datasets, and through a reduction in current barriers (ethical, regulatory).[23] There is still a long way to go to achieve this; even with a pandemic, the willingness to share IPD is low.[24]

Future prospective coordination and collaboration for more consistent data collection will make the efforts of IPD meta-analysis more beneficial to the clinical community.[18] IPD meta-analysis is an ethical, efficient research method that reuses data and enables investigations of participant-level characteristics, treatment effect and outcomes that would otherwise not be feasible with most current trial sizes. They generate additional knowledge from existing studies, thereby reducing the need for more primary studies and limiting research waste. The clinical utility of IPD meta-analysis depends upon well-conducted and planned trials, the collection of a set of minimum variables, and ethics and informed consent in primary trials that allow for future data sharing.

Competing interests: Nil.

Source(s) of Support: Nil.

Acknowledgements: We thank Rachel Ogilvie and Heather Taylor for helpful comments and suggestions on this Research Note.

Provenance: Invited. Peer reviewed.

Correspondence: Jill A Hayden, Community Health & Epidemiology, Dalhousie University, Canada. Email: jhayden@dal.ca

Jill A Hayden[a] and Richard D Riley[b]

[a]*Community Health & Epidemiology, Dalhousie University, Canada*
[b]*Centre for Prognosis Research, School of Medicine, Keele University, United Kingdom*

References

1. Higgins JPT, et al, eds. *Cochrane handbook for systematic reviews of interventions.* Cochrane; 2020. Version 6.1.
2. Riley RD, et al. *BMJ.* 2010;340:221–228.
3. Riley RD, et al, eds. *Individual participant data meta-analysis: A handbook for healthcare research.* Chichester: Wiley; 2021.
4. Tierney JF, et al. *PLoS Med.* 2015;12:e1001855.
5. Burke DL, et al. *Stat Med.* 2017;36:855–875.

6. Riley RD, et al. *Stat Med*. 2020;39:2115–2137.
7. Stewart LA, et al. *JAMA*. 2015;313:1657–1665.
8. Levis B, et al. *JAMA*. 2020;323:2290–2300.
9. Holden S, et al. *Pain*. 2021. https://doi.org/10.1097/j.pain.0000000000002184. Online only.
10. Hudda MT, et al. *BMJ*. 2019;366:4293–4303.
11. Hayden JA, et al. *Br J Sports Med*. 2019;54:1277–1293.
12. Saragiotto BT, et al. *J Clin Epidemiol*. 2016;79:3–9.
13. Gurung T, et al. *Physiotherapy*. 2015;101:243–251.
14. Mistry D, et al. *Spine*. 2014;39:618–629.
15. Tierney JF, et al. *PLoS Med*. 2020;17:e1003019.
16. Riley RD, et al. *Res Synth Methods*. 2015;6:157–174.
17. Riley RD, et al. *BMJ*. 2017;358:j3932.
18. Seidler AL, et al. *BMJ*. 2019;367:l5342.
19. Ventresca M, et al. *BMC Med Res Methodol*. 2020;20:113–131.
20. Ahmed I, et al. *BMJ*. 2012:344:d7762.
21. Stewart LA, Clarke MJ. *Stat Med*. 1995;14:2057–2079.
22. Nevitt SJ, et al. *BMJ*. 2017;357:j1390.
23. Scutt P, et al. *BMJ Open*. 2020;10:e038765.
24. Li R, et al. *Trials*. 2021;22:153–157.