# 1

# Individual Participant Data Meta-Analysis for Healthcare Research

*Richard D. Riley, Lesley A. Stewart, and Jayne F. Tierney*

COPYRIGHTED MATERIAL

## 1.1 Introduction

Healthcare and clinical decision-making should be guided by the evidence arising from high-quality research studies. Often a single study is insufficient to make firm recommendations, and so multiple studies are conducted to address the same research question. This motivates the need for *evidence synthesis*: the combination of data from multiple studies to provide an overall summary of current knowledge. For example, when multiple randomised trials have examined the effect of a particular treatment, evidence syntheses are needed to combine and summarise the information from these trials, in order to establish whether the treatment is effective or not.

*Systematic reviews* are the cornerstone of evidence synthesis and evidence-based decision-making in healthcare. They use transparent methods to identify, appraise and combine a body of research evidence, with the goal of producing summary results that guide best practices for stakeholders including patients, clinicians, health professionals, and policy-makers. Systematic review methodology has been championed by organisations such as Cochrane, who publish systematic reviews in the Cochrane Library summarising the effects of interventions,[1] the accuracy of diagnostic tests,[2] the prognostic effect of particular factors,[3] and the performance of risk prediction models.[4] Most systematic reviews include a *meta-analysis*,[5] which is a statistical technique for combining (synthesising) quantitative data obtained from multiple research studies. Traditionally, most meta-analyses have used aggregate data extracted from study publications, but there is growing demand for meta-analyses that utilise individual participant data (IPD).[6–9]

This book is intended as a comprehensive handbook for healthcare researchers undertaking IPD meta-analysis projects. In this introductory chapter, we clarify differences between IPD and aggregate data, and outline why IPD meta-analysis projects are increasingly needed. Then, we detail the scope of our book and its intended audience, and signpost where to find material in subsequent chapters.

## 1.2 What Is IPD and How Does It Differ from Aggregate Data?

IPD refers to the raw information recorded for each participant in a research study (e.g. a randomised trial), such as baseline characteristics, prognostic factors, treatments received, outcomes and follow-up details, and can be represented by a dataset containing a separate row per participant and columns containing values for each participant-level variable. For example, IPD for a randomised trial of anti-hypertensive treatment will usually include the pre- and post-treatment blood pressure level, a treatment group indicator, important clinical characteristics and prognostic factors

*Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research*, First Edition.
Edited by Richard D. Riley, Jayne F. Tierney, and Lesley A. Stewart.
© 2021 John Wiley & Sons Ltd. Published 2021 by John Wiley & Sons Ltd.

recorded at baseline (such as age, sex, BMI and comorbidities), and relevant follow-up information (such as time to cardiovascular disease or death). An IPD meta-analysis project, therefore, involves the collection, checking, harmonisation and synthesis of IPD from multiple studies to answer particular research questions. An excerpt of IPD collected from 10 randomised trials for an IPD meta-analysis project is given in Box 1.1(a), after harmonisation into a single dataset ready for meta-analysis to summarise the effect of anti-hypertensive treatment. This dataset contains a single row *per participant* in every trial.

In contrast, aggregate data refers to information averaged or estimated across all participants in a particular study, such as the treatment effect estimate, the total participants, and the mean age and proportion of males in each treatment group. Such aggregate data are derived from the IPD, and therefore the IPD can be considered the original source material. A conventional meta-analysis uses aggregate data (e.g. as extracted from study publications), rather than IPD. An example of aggregate data obtained from 10 randomised trials of anti-hypertensive treatment is shown in Box 1.1(b), after collation into a single dataset ready for meta-analysis. This dataset contains a single row *per trial*.

## 1.3 IPD Meta-Analysis: A New Era for Evidence Synthesis

*"Data sharing is an important part of ensuring trust in research, and it should be the norm."*[10]

IPD meta-analysis projects began to emerge in the late 1980s and early 1990s,[11,12] originating mainly in the cancer and cardiovascular disease fields.[13] Calls to support IPD meta-analysis grew strongly throughout the 1990s alongside the formation of methodology working groups,[8,14] in particular the Cochrane IPD Meta-Analysis Methods Group (https://methods.cochrane.org/ipdma/).[7] In the decades since, the number of IPD meta-analysis projects has risen sharply (Figure 1.1). Early meta-analyses based on IPD were commonly described as *overviews* or *pooled analyses,*[7,11,15,16] until *IPD meta-analysis* emerged as the preferred, and now most widely used, label. The IPD abbreviation initially referred to *individual patient data*, but now *individual participant data* is the more inclusive and accepted term.

The growth of IPD meta-analysis projects reflects their potential to revolutionise healthcare research,[14,17] especially as they align with three major contemporary initiatives: *reducing research waste*,[18] *data sharing*,[19–24] and *personalised healthcare*.[25,26] The sharing of IPD maximises the contribution of existing data from millions of research participants, and so is becoming an increasingly frequent stipulation of research funding. Leading medical journals now require data-sharing statements, with some even enforcing the sharing of IPD on request.[23] This has led to dedicated data-sharing platforms and repositories being established to house IPD from existing studies.[27–31] Furthermore, as the drive for personalised healthcare (also known as stratified or precision medicine) continues,[25,26] researchers have recognised that, compared to using published aggregate data, IPD allows a more reliable evaluation of how participant-level characteristics are associated with outcome risk and response to treatment.[32,33] Thus, IPD meta-analysis projects are now central to modern evidence synthesis in healthcare.

## 1.4 Scope of This Book and Intended Audience

*Meta-Analysis Using Individual Participant Data: A Handbook for Healthcare Research* provides a comprehensive introduction to the fundamental principles and methods that healthcare

**Box 1.1 Example of individual participant data (IPD) and how it differs from aggregate data**

**Illustrative example of 10 randomised trials examining the effect of anti-hypertensive treatment**

(a) IPD

- The following table shows hypothetical IPD collected, checked and harmonised from 10 randomised trials examining the effect of anti-hypertensive treatment versus control in participants with hypertension.
- Each row provides the information for each participant in each trial, and each column provides participant-level information such as baseline characteristics and outcome values.
- Only a subset of the IPD is shown for brevity, as in reality many more rows and columns will be needed for each trial, to include all available participants and variables.

| Trial ID | Participant ID | Treatment group, 1 = treatment 0 = control | Age (years) | SBP before treatment (mmHg) | SBP at 1 year (mmHg) |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 46 | 137 | 111 |
| 1 | 2 | 1 | 35 | 143 | 133 |
| | | (other rows for trial 1 omitted for brevity) | | | |
| 1 | 1454 | 0 | 62 | 209 | 219 |
| 2 | 1 | 0 | 55 | 170 | 155 |
| 2 | 2 | 1 | 38 | 144 | 139 |
| | | (other rows for trial 2 omitted for brevity) | | | |
| 2 | 337 | 1 | 44 | 153 | 129 |
| | | (rows for trials 3 to 9 omitted for brevity) | | | |
| 10 | 1 | 0 | 71 | 149 | 128 |
| 10 | 2 | 1 | 59 | 168 | 169 |
| | | (other rows for trial 10 omitted for brevity) | | | |
| 10 | 4695 | 0 | 63 | 174 | 128 |

- This IPD can be used to produce aggregate data for each trial, as shown in the table on the following page.

*(Continued)*

**Box 1.1 (Continued)**

(b) Aggregate data

- Now each row corresponds to a particular trial, and each column is a trial-level variable containing aggregated data values such as the total number of particulars and the mean age in each group.

| Trial ID | Number of participants | | Mean age (years) | | Mean SBP before treatment (mmHg) | | Mean SBP at 1 year (mmHg) | | Treatment effect on SBP at 1 year adjusted for baseline (treatment minus control) Estimate (variance) |
|---|---|---|---|---|---|---|---|---|---|
| | Control | Treatment | Control | Treatment | Control | Treatment | Control | Treatment | |
| 1 | 750 | 704 | 42.36 | 42.17 | 153.05 | 153.88 | 139.75 | 132.54 | –6.53 (0.75) |
| 2 | 199 | 138 | 69.57 | 69.71 | 191.55 | 188.30 | 179.89 | 164.67 | –13.81 (4.95) |
| | | | | (rows for trials 3 to 9 omitted for brevity) | | | | | |
| 10 | 2297 | 2398 | 70.21 | 70.26 | 173.94 | 173.75 | 165.24 | 154.87 | –10.26 (0.20) |

*Source:* Richard Riley.

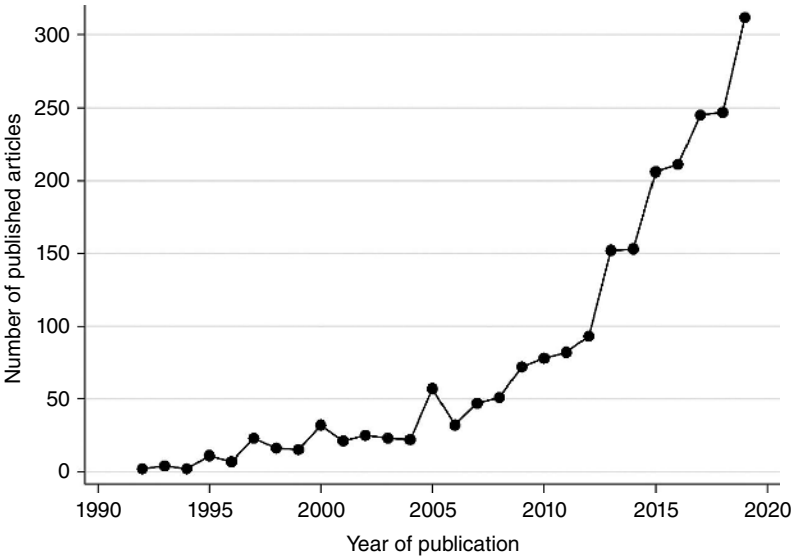


**Figure 1.1** Number of published IPD meta-analysis articles over time, based on a crude search* in PubMed
*Source:* Richard Riley.
*from searching for the following keywords in the Title or Abstract of the article: (meta-analysis AND individual patient data) OR (meta-analysis AND individual participant data) OR (meta-analysis AND IPD).

researchers need when considering, conducting or using IPD meta-analysis projects. Written and edited by researchers with substantial experience in the field, the book details key concepts and practical guidance alongside illustrated examples and summary learning points.

IPD meta-analysis projects require a multi-disciplinary research team, including clinicians and healthcare professionals, statisticians, evidence synthesis experts, search and information specialists, database managers, trialists, and patient and public advisory groups, amongst others. Therefore, this book is aimed at a broad audience, and guides the reader through the journey from initiating and planning IPD projects to obtaining, checking, and meta-analysing IPD, and appraising and reporting findings. Very little prior knowledge is required. We assume readers are aware of the importance of systematic reviews and meta-analysis in general, and are reading this book to help guide their decisions as to whether to take the IPD approach; to learn what an IPD project entails (from start to finish); and to understand appropriate methodology and best practice, for example to inform protocols, data retrieval plans, statistical analyses, bias assessments, reporting standards, and critical appraisal.

Our book is split into five parts. Parts 1 to 3 focus on the synthesis of IPD from randomised trials to examine treatment effects. Parts 4 and 5 branch out to cover special topics and applications, including diagnosis, prognosis and prediction. **Part 1** includes chapters 2 to 4, and covers practical guidance for initiating, planning and conducting IPD meta-analysis projects. **Part 2** includes chapters 5 to 8, and covers fundamental two-stage and one-stage statistical methods for conducting an IPD meta-analysis of randomised trials to examine a treatment effect. These chapters are more technical than others, but should still be broadly accessible, as recommendations and illustrated examples are given throughout to reinforce the key messages. **Part 3** includes Chapters 9 to 11, and focuses on the critical appraisal and dissemination of IPD projects. **Part 4** includes Chapters 12 to 14, and

covers special topics in statistics, including calculating power (in advance of IPD collection) and analysing multiple outcomes and multiple treatments. **Part 5** concludes with Chapters 15 to 18, which broaden application of IPD projects to the evaluation of diagnostic tests, prognostic factors, and clinical prediction models.

This book is the first to be devoted entirely to IPD meta-analysis projects, and complements other textbooks on systematic reviews and meta-analysis that focus mainly on the aggregate data approach, such as the following.[1,34–37] A general statistical textbook would also provide complementary reading to Part 2 of this book.[38–41] Relevant methods for IPD meta-analysis of prognosis studies are introduced in *Prognosis Research in Healthcare: Concepts, Methods and Impact*,[32] and Part 5 builds extensively on this work. Detailed information is also available on our companion website for this book: www.ipdma.co.uk. Introductory videos are included, alongside links to relevant publications, talks, training courses, and workshops. Statistical code is also provided for educational purposes, so that readers can replicate various examples given throughout the book and reinforce their learning.