Sharing data–taming the beast: barriers to meta-analyses of individual patient data (IPD) and solutions

Marienke van Middelkoop [iD],[1] Stefan Lohmander [iD],[2] Sita M A Bierma-Zeinstra[1]

Systematic reviews aim to collate all empirical evidence that fits prespecified eligibility criteria to answer a specific research question. Some systematic reviews undertake a meta-analysis to statistically combine study results and provide a more precise estimate of treatment effects. These meta-analyses are commonly based on aggregate data, extracted from publications or obtained from the original authors of these papers[1] but aggregating data limits the options for in depth analysis.

[1]Department of General Practice, Erasmus MC Medical University Center, Rotterdam, The Netherlands
[2]Faculty of Medicine, Department of Clinical Sciences Lund, Orthopaedics, Lunds Universitet, Lund, Sweden

Correspondence to Dr Marienke van Middelkoop, General Practice, Erasmus Medical Center, Rotterdam 3015 GD, The Netherlands; m.vanmiddelkoop@erasmusmc.nl
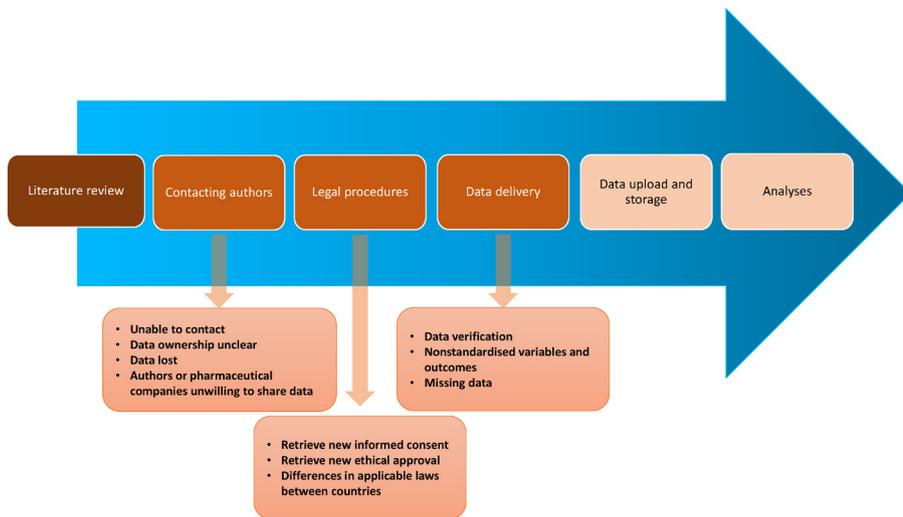
Figure 1 Steps to be taken in individual patient data approach and potential barriers and challenges.

Table 1	Critical elements to improve future individual patient data analyses
1.	Change our mind-set: to openly share data is a win-win situation
2.	Collect and report minimum core outcome trial data following standards for conditions of interest
3.	Store annotated data for both academic and industry sponsored trials in open access repositories
4.	Harmonise ethics and legal issues for data-access and reuse
5.	Provide funding and guarantees for sustainable open data repositories

Meta-analysis of individual patient data (IPD) appeared in the 1990s[2] to address these problems. A great advantage of IPD analysis is that it offers investigators the opportunity to investigate whether an intervention is differentially effective for different types of participants. Quantifying interaction effects using IPD increases power and generalisability of results and is considered the gold standard for subgroup analyses.[3]

We congratulate Hayden and colleagues (linked paper *BJSM* 2020, bjsports-2019-101205)[4] who undertook an IPD to identify subgroups of patients who particularly benefit from exercise therapy for low back pain.[4] From a total pool of 56 eligible trials, the authors retrieved data from 27 studies (3514 participants). This reflects the major challenge when performing IPD analyses—retrieving raw data from multiple trials.

Let us drill down on some specific barriers to successfully obtaining the 'D' in IPD—data (figure 1). The first step, contacting original authors, can be challenging and some authors reached do not respond.[5][6] Moreover, authors often refuse to participate or report that data is no longer available at their institute, or that they do not have intellectual property rights to the original data. The latter is particularly common when pharmaceutical companies own the data.[5–7]

Even when authors are willing to share data, the taming of the beast is just beginning. Data regulations and governing laws can be quite complex. As they differ markedly between countries, and most IPD meta-analyses include data from many countries, it can be extremely difficult to obtain a data delivery agreement signed by all parties. In some countries such as Canada and Australia, new analyses (ie, the IPD meta-analysis) require new ethics committee approval.

Once researchers have retrieved data, they face further challenges. Hayden and colleagues tried to verify the data and were able to replicate the main study outcomes of fewer than 50% of the trials. The authors were able to analyse 18 potential effect modifiers but many of these had up to 75% missing data (eg, 'history of low back pain'). Potential effect modifiers are often not measured in the foundation studies or are inconsistently available—this greatly constrains researchers from analysing potential treatment effect modification and is a frequent problem in IPD analyses.

SYSTEM-WIDE EFFORTS TO OVERCOME SOME OF THESE BARRIERS

What is being done to stimulate open access to research data? Funding agencies increasingly require data be shared after a project is finished and in 2016 the Council of the European Union encouraged member countries to transition to an open science system. In the Netherlands, ZonMW (The Netherlands Organisation for Health Research and Development) supports FAIR (findable, accessible, interoperable, reusable research data), requiring researchers to share their data to contribute to future research.[8] The US National Institutes of Health has a similar requirement.

Scientific journals increasingly encourage open access of data. *BMJ* has adopted different policies on data sharing, depending on the specific journal. These include the requirement and encouragement to openly and publicly make generated data available on publication. Notwithstanding, this is still voluntary in many of these journals. *BMJ* also accepts DataCite DOIs that make it possible to cite publicly available used data in reference lists.

Successful examples of studies with open data by design include the osteoarthritis (OA) initiative and the cohort hip and cohort knee (CHECK)-studies, two multicentre, longitudinal, prospective observational studies of knee and hip OA.[9][10] All collected individual data are openly accessible or available on application, resulting in more than 600 publications, attesting to the power of open data sharing.

Multiple initiatives have now been launched to build collaborations for the development of IPD banks to facilitate data accessibility, such as the OA Trial Bank for clinical OA research and the World COACH study for morphological data of the hip.[11] These initiatives provide consistent and transparent rules of collaboration and agreements for sustainability and accessible sharing of data.

Sport Data Valley in the Netherlands aims to connect sport with science, government and companies. All sport science and medicine related data can be uploaded into the repository and access rights are adjustable per dataset, and data ownership remains at the principal investigator. Such repositories make data widely accessible to a broad audience.[12] Other examples of controlled access repositories of data include the clinicalstudydatarequest.com and the Yale University Open Data Access (YODA) project.

Although many challenges remain, the time investment and barriers facing IPD analyses should decrease in the coming years. We expect that data will be richer and more consistent given the disease-specific reporting standards and core data sets launched in many fields of research. However, researchers and pharmaceutical companies must be willing to share data so that the potential value of IPD analyses is realised. To maximise the use of individual participant data collected in clinical studies is also to fulfil the ethics contract with the study participants (table 1).

Twitter Marienke van Middelkoop @mvanmiddelkoop

OPEN ACCESS

Check for updates

Linked

► http://dx.doi.org/10.1136/bjsports-2019-101205

ORCID iDs
Marienke van Middelkoop http://orcid.org/0000-0001-6926-0618
Stefan Lohmander http://orcid.org/0000-0002-5424-9448

REFERENCES

1. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011. Available: www.handbook.cochrane.org
2. Ioannidis J. Next-Generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *Br J Sports Med* 2017;51:1456–8.
3. Groenwold RHH, Donders ART, van der Heijden GJMG, *et al*. Confounding of subgroup analyses in randomized data. *Arch Intern Med* 2009;169:1532–4.
4. Hayden JA, Wilson MN, Stewart S, *et al*. Exercise treatment effect modifiers in persistent low back pain: an individual participant data meta-analysis of 3514 participants from 27 randomised controlled trials. *Br J Sports Med* 2019. doi:10.1136/bjsports-2019-101205. [Epub ahead of print: 28 Nov 2019].
5. van Middelkoop M, Arden NK, Atchia I, *et al*. The oa trial bank: meta-analysis of individual patient data from knee and hip osteoarthritis trials show that patients with severe pain exhibit greater benefit from intra-articular glucocorticoids. *Osteoarthritis Cartilage* 2016;24:1143–52.
6. Runhaar J, Rozendaal RM, van Middelkoop M, *et al*. Subgroup analyses of the effectiveness of oral glucosamine for knee and hip osteoarthritis: a systematic review and individual patient data meta-analysis from the oa trial bank. *Ann Rheum Dis* 2017;76:1862–9.
7. Fleetcroft R, Ford J, Gollop ND, *et al*. Difficulty accessing data from randomised trials of drugs for heart failure: a call for action. *BMJ* 2015;351:h5002.
8. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
9. NIH. Osteoarthritis initiative: NIH, 2019. Available: https://www.niams.nih.gov/grants-funding/funded-research/osteoarthritis-initiative
10. Utrecht UMC. The check study documentation and data: University medical center Utrecht, 2020. Available: https://www.check-onderzoek.nl/
11. van Middelkoop M, Dziedzic KS, Doherty M, *et al*. Individual patient data meta-analysis of trials investigating the effectiveness of intra-articular glucocorticoid injections in patients with knee or hip osteoarthritis: an oa trial bank protocol for a systematic review. *Syst Rev* 2013;2:54.
12. ZonMW. Sport data Valley: ZonMW, 2020. Available: https://www.sportinnovator.nl/sport-data-valley/