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

Marienke van Middelkoop,1 Stefan Lohmander,2 Sita M A Bierma-Zeinstra1

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 papers1 but aggregating data limits the options for in depth analysis.

1Department of General Practice, Erasmus MC Medical University Center, Rotterdam, The Netherlands
2Faculty 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@erasusmc.nl
Meta-analysis of individual patient data (IPD) appeared in the 1990s 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.

We congratulate Hayden and colleagues (linked paper BJSM 2020, bjsports-2019-101205) who undertook an IPD to identify subgroups of patients who particularly benefit from exercise therapy for low back pain. 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. 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.

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. 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. 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. 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. 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

Contributors Work was initially conceived by MvM, SL, SMAB-Z. Substantial contributions to the conception of the work were made by all authors. Drafting and revising the work critically was done by all authors. Final approval of the version published was given by revising the work critically was done by all authors.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.


Accepted 21 January 2020
Published Online First 30 January 2020

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

doi:10.1136/bjsports-2019-101892

ORCID iDs

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

REFERENCES


