BMJ Open

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-049228 |
| Article Type: | Original research |
| Date Submitted by the Author: | 20-Jan-2021 |
| Complete List of Authors: | Ohmann, Christian; European Clinical Research Infrastructure Network (ECRIN),<br>Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre<br>Siebert, Maximilian; University Rennes, CHU Rennes, CIC 1414 (Centre d'Investigation Clinique de Rennes)<br>Motschall, Edith ; University of Freiburg Faculty of Medicine, Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg,<br>Naudet, Florian; University Rennes, CHU Rennes, INSERM CIC 1414 (Centre d'Investigation Clinique de Rennes) |
| Keywords: | Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Information management < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review

**(Date: 18.01.2021)**

**Authors**

C. Ohmann, ORCID: 0000-0002-5919-1003, christianohmann@outlook.de
D. Moher, ORCID: 0000-0003-2434-4206, dmoher@ohri.ca
M. Siebert, ORCID: 0000-0003-4385-5773, maximiliansiebert91@gmail.com
E. Motschall, ORCID: 0000-0001-7795-242X, motschall@imbi.uni-freiburg.de
F. Naudet, ORCID: 0000-0003-3760-3801, floriannaudet@gmail.com

**Corresponding author:**

Prof. Dr. Christian Ohmann
European Clinical Research Infrastructure Network (ECRIN)
Kaiserswerther Str. 70
40477 Düsseldorf
Germany
Tel.: +49 211 5142490
Mobile: + 49 16097945385
Email: christianohmann@outlook.de

**Co-authors :**

David Moher
Ottawa Hospital Research Institute
Ottawa Methods Centre
Ottawa, On
Canada

Maximilian Siebert, Florian Naudet
Université de Rennes 1
CHU Rennes
INSERM, CIC 1414 (Centre d'Investigation Clinique de Rennes)
Rennes, Bretagne
France

Edith Motschall
Institute of Medical Biometry and Statistics
Faculty of Medicine and Medical Center
University of Freiburg
Germany

**Key words:**

clinical trial, individual participant data, data sharing, scoping review, impact

**Word count:**

11929

## ABSTRACT

**Objectives**

To explore the impact of data sharing initiatives on the intent to share data, to actual share data, to use shared data and to provide research output and impact from shared data

**Eligibility criteria**

All studies, investigating data sharing practices with respect to individual participant data (IPD) from clinical trials.

**Sources of evidence**

We searched Medline databases, the Cochrane Library, Science Citation Index Expanded and Social Sciences Citation Index via Web of Science, preprints and proceedings of the International Congress on Peer Review and Scientific Publication. In addition, we inspected major clinical trial data-sharing platforms, contacted major journals/publishers, editorial groups and some funders.

**Charting methods**

Two reviewers independently extracted information on methods and results of identified resources with a standardised questionnaire. A map of the extracted data was constructed and accompanied by a narrative summary for each outcome domain.

**Results**

93 studies identified through the literature search and 5 additional information sources were included in the scoping review. Most studies were descriptive and focused on early phases of the data sharing pipeline. While the willingness to share IPD from clinical trials is extremely high, the actual data sharing rates are suboptimal. Survey of journal data suggests a poor to moderate enforcement of the policies by publishers. Metrics from platforms suggest that a large majority of data remain unrequested. When requested, the purpose of the re-use is more often secondary analyses and meta-analyses, rarely re-analyses. Finally, studies focused on real impact of data sharing were rare and used surrogates such as citation metrics.

**Conclusions**

There is currently a gap in the evidence base investigating the impact of IPD sharing, which causes uncertainties in the implementation of current data-sharing policies. High level evidence is needed to assess whether the value of medical research increases with data sharing practices.

**Strengths and limitations of this study**

- Exhaustive review of both the literature and the most important initiatives in data sharing

- Analysis of the full data sharing pipeline covering intention to share, actual sharing, use of shared data, research output and impact

- Retrieving and synthetizing information proved to be difficult because of a very siloed landscape where each initiative/platform operates with its own metrics

- Data sharing is a moving target in a rapidly changing environment with more and more new initiatives.

- Only a limited research output from data sharing is available so far

**Funding**

No specific funding for this review.
FN work on data sharing is supported by a grant from the French National Research Agency – ANR (Reproducibility in Therapeutic Research / ReITheR: ANR-17-CE36-0010-01).
CO work on data sharing is supported by funding from the European Union's Horizon 2020 Research and Innovation Programme (CORBEL, under grant agreement n° 654248).
DM is supported by a Ottawa University Research Chair (grant number: N/A)..

**Competing interests**

None of the authors have any competing interests.

**Author's contribution**

CO, DM and FN developed the study protocol. Search strategy was developed and implemented by EM. Selection of source of evidence and assessment was performed by CO and FN. Contact with initiatives/platforms/journals/publishers was performed by MS. In case of disagreements, these were resolved by consensus and, when necessary, in consultation with DM. The first draft of the manuscript was written by CO and FN. All others revised and approved the final manuscript.

**Data sharing statement**

All data relevant to the study are included in the article or uploaded as supplementary Information.

## INTRODUCTION

### Rationale

Data sharing is increasingly being recognized as a key requirement in clinical research.[1] Within clinical research, data sharing can enhance reproducibility and the generation of new knowledge, but it also has an ethical and economical dimension.[2] Scientifically, sharing makes it possible to compare or combine the data from different studies, and to more easily aggregate it for meta-analysis. It allows conclusions to be re-examined and verified or, occasionally, corrected, and it can allow new hypotheses to be tested. Sharing can therefore increase data validity, but it also squeezes more value from the original research investment, as well as helps to avoid unnecessary repetition of studies. Agencies and funders are referring more and more to the economic advantages of data reuse. Ethically, data sharing provides a better way to honour the generosity of clinical trial participants, because it increases the utility of the data they provide. Despite the high potential of sharing clinical trial data, the launch and implementation of several data sharing initiatives and platforms and outstanding examples related to the value of data sharing,[3] till now data sharing is not the norm in clinical research compared to many other scientific disciplines.[4] One major hurdle is that clinical trial data are about individuals and their health status, and as such requires specific measures to protect their privacy.

To support data sharing of IPD in clinical trials, several organisations have developed generic principles, guidance and practical recommendations for implementation. In 2016, the International Committee of Medical Journal Editors (ICMJE), a small group of medical journal editors, published an editorial[5] stating that "it is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk". The ICMJE considers that there is an implicit social contract imposing an ethical obligation that the results lead to the greatest possible benefit to society. The ICJME proposed to require that de-identified individual patient data (IPD) are made publicly available no later than 6 months after publication of the main trial results. Such a time period would be useless for public health emergencies like COVID 19. However, the ICMJE proposal triggered debate and a large number of trialists were reluctant to adopt this new norm[6] regarding the feasibility of the proposed requirements, the necessary resources, the real or perceived risks to trial participants, and the need to protect the interests of patients and researchers.[7]

Despite the cultural change towards sharing clinical trial data and the major commitment of scientific organisations, funders and initiatives, overall, there is still a lack of effective policies in the biomedical literature to ensure that underlying data is maximally available and reusable. The only requirement appears to be a data management plan or a data sharing plan. A few journals require data sharing and, for those who require data sharing, guidelines are heterogeneous and somewhat ambiguous.[8] Nevertheless, some innovative and progressive funders (e.g. Wellcome Trust, Bill & Melinda Gates Foundation), and publishers/journals (e.g. PLOS, The BMJ) had adopted strong data sharing policies. As part of a wider cultural shift towards more open science, there have been various attempts to explore how clinical researchers can best plan for data sharing and prepare their 'raw' IPD so that it becomes available to others[9] – albeit often under controlled access conditions rather than simply being publicly available on-line[10] and structure that data to make it FAIR (findable, accessible, interoperable and reusable).[11] Meanwhile several data sharing platforms and repositories are available and in use to practically support the data sharing process in clinical research (e.g. YODA,CDSR, vivli, NIH). A considerable number of individual studies has been performed to access and explore data sharing from clinical trials under different circumstances and within different frameworks. What is strongly needed is a scoping review helping to get an overview on the status of implementation of data sharing as a whole and implications originating from the available evidence.

### Objectives

In this scoping review we explored the impact of data sharing initiatives on the intent to share data, status of data sharing, use of shared data and impact of research outputs from shared data.

## METHODS

### Protocol and registration

The protocol of the study was registered on the Open Science Framework on September the 12th 2018 (registration number: osf.io/pb8cj). The protocol followed the methodology manual published by the Joanna Briggs Institute for scoping reviews.[12] Methods and results have been reported using the PRISMA extension for scoping reviews (PRISMA-ScR).[13]

**Eligibility criteria**

The following eligibility criteria for studies were used:

All study designs were eligible, including case studies, surveys, metrics and experimental studies, utilizing qualitative or quantitative methods. Only reports (published or unpublished) in English, German, French or Spanish were considered.

We included all studies and reports 1/ providing information on current data sharing practices of IPD for clinical trials and 2/ reporting on one or more of five outcome domains defined according to the data sharing pipeline presented in **Box 1**.

**1. Intention to data sharing**

There is an intention to share data, formulated by a stakeholder (e.g. sponsor/PI, funder). This can be done by a written data sharing commitment or by a declaration included in the trial registration. This also includes surveys on attitudes towards data sharing.

**2. Actual data sharing**

Data are truly made available for data sharing to secondary users. This is important because there are cases known where the data are offered for sharing but sharing does not take place according to a possible hidden agenda and changing of plans.

**3. Use of shared data**

Shared data may be used for various purposes. It may be used as background for research, usually not leading to research outputs. This covers use for education, researcher training and data understanding. Studies that should lead to new research outputs include 1/ validation/reproducibility of results, 2/ further additional analyses (prognostic models, decision-support, subgroup analyses, etc.) and 3/ IPD meta-analyses.

**4. Research outputs from shared data**

Research outputs are scientific presentations, reports and publications.

**5. Impact of research output from shared data**

Research output from shared data may have an impact on medical research (e.g. development of new hypotheses and methods) and/or medical health (e.g. changing treatment via guidelines).

**Box 1:** Definitions used for the 5 outcome domains

In the scoping review only data sharing of IPD from clinical trials was considered. We defined clinical trials following the clinicaltrials.gov definition as "a clinical study is a research study involving human volunteers (also called participants) that is intended to add to medical knowledge. There are two types of clinical studies: interventional studies (also called clinical trials) and observational studies. Clinical trial is another name for an interventional study."[14] We therefore considered any interventional clinical studies (no matter whether they were randomised or not) and we didn't consider studies on data sharing concerning observational and non-clinical studies (e.g. genomics) nor different fields outside medicine (e.g. economics).
We included studies that investigated and reported information on current data sharing practices performed without restrictions in terms of promotional initiatives, type of repository or platform (see **Box 2** for

definitions) that promote data sharing practices (e.g. at the editor level, at the funder level, at the scientist level etc.). We considered many different types of studies (e.g. experimental studies, surveys, metrics, quality assurance studies, qualitative research, reviews, reports) as the inclusion criteria were not methods-specific but rather content-specific**.**

| |
|---|
| **initiative**<br><br>Major activities of an organization (or a network of several organizations) to actively promote data sharing in this area (e.g. PHRMA/EFPIA, Nordic Trial Alliance, IOM, ICMJE, RDA).<br><br>**Repository**<br><br>Large database infrastructures set up to manage, share, access and archive researchers' datasets from clinical trials. Repositories may be specialised and dedicated to specific disciplines (e.g. FreeBird, BioLINCC) or more general (e.g. FigShare, Dryad).<br><br>**Platform**<br><br>A computer environment where researchers can find datasets from clinical trials across different repositories and where additional functionalities (e.g. protected analysis environment) are provided (e.g. CSDR, YODA, project Datasphere, Github). |

**Box 2:** Definitions used for initiative, repository and platform

**Information sources**

The identification of studies was performed in two complementary ways:

a) A systematic literature search in bibliographic databases (MEDLINE databases, Cochrane Library, Science Citation Index Expanded and Social Science Citation Index). In addition, a preprint server and proceedings were searched
b) Inspection and eventually contacts of known information sources (e.g. webpages, documents and reports from platforms, funder, publisher) to explore whether they had an evaluation component and provided detailed research output from shared data (see supplementary material 1).

Between 25/01/2019 and 12/06/2019 (with an update in 02/11/2020), one researcher (MS) inspected (and when necessary contacted) major clinical trials data-sharing platforms to explore whether they had an evaluation component and provided detailed of research output from shared data (see **Supplementary Material 1**). Similarly, during the same time period, the researcher contacted major journals and/or publishers and/or editorial groups (The BMJ, PLOS, The Annals of Internal Medicine, BioMedCentral (Springer/Nature), F1000Research. Some funders (see **Supplementary Material 1**) were also contacted as well as preprints repositories (bioRxiv, PeerJ, Preprints.org, PsyArXiv and MedRxiv. For sake of completeness, he has also contacted ASAPbio (Accelerating Science and Publication in biology) and the Center for Open Science for the same information as well as three International Congress on Peer Review and Scientific Publication conference abstracts. In addition, when relevant references were found in various papers these references were included (snowballing searches).

**Search**

On 29/10/2018 (update on 12/09/2020), a researcher (EM) searched the Medline databases for indexed and non-indexed citations via Ovid from Wolters Kluwer, the Cochrane Library via Wiley, Science Citation Index Expanded and Social Sciences Citation Index via Web of Science from Clarivate Analytics for articles meeting our inclusion criteria.

The detailed search terms for the MEDLINE databases, the Cochrane Library and the Web of Science databases can be found in **Supplementary Material 2**. The main search strategy developed by CO, DM und FN was peer reviewed independently (by a senior medical documentalist, EM who joined the team subsequently) using evidence-based guidelines for Peer Review of Electronic Search Strategies (PRESS).[15] Discrepancies were resolved between the authors and EM performed the search. All references were managed and de-duplicated using a reference manager system (Endnote).

On the 23/01/2019 (update in 02/11/2020), two researchers (MS and FN) independently searched for relevant pre-prints through OSF PREPRINTS using the search function to find all papers relevant to medicine with the following keyword (trial* OR random*). On the 29/01/2019, the two researchers independently searched the proceedings of the three latest International Congress on Peer Review and Scientific Publication for relevant abstracts (2009, 2013 and 2017).

**Selection of sources of evidence**

Selection of source of evidence was performed by two independent reviewers (CO and FN). Contact with initiatives/platforms/journals/publishers was performed by a single reviewer (MS). In case of disagreements, these were resolved by consensus between CO and FN and, when necessary, in consultation with a third reviewer (DM).

**Data charting process**

We developed a data collection form and pilot tested it on 10 randomly selected research papers that were later included in our final study. In case of disagreements, these were resolved by consensus and, when necessary, in consultation with a third reviewer (DM).

**Data items**

For each research paper included according to the selection criteria we extracted: 1/ basic information on the paper (type of study exploring data sharing practices, authors, year, references, and type of initiative and/or repository and/or platform studied), 2/ information on the material shared (sharing of data, code, programs and material), 3/ whether it reported data about one or more of five outcomes domains defined box 1, 4/ how were these outcome domains assessed, and 5/ we described qualitatively the main results observed on these outcomes.

For each of the data-sharing platforms, publishers and funders providing detailed research output from shared data, we extracted the following information (authors, date of request, date of publication, type of re-use). We initially planned to describe importance of the re-uses in qualitative terms and the observed results of the re-use (i.e. “positive” or “negative” study) but these two characteristics were difficult to extract with very poor inter-rater agreement and we decided not to detail them.

**Critical appraisal of individual sources of evidence**

Included studies were classified according to study type (e.g. survey, metrics, experimental). Potentially relevant characteristics of included studies with regard to their internal-external validity were not assessed systematically and with a specific tool but explored when one of the two reviewers judged it relevant and were discussed thoroughly for each study between the reviewers.

**Synthesis of results**

No outcome was prioritized since there was no quantitative synthesis for this study. All outcomes were described separately in sections corresponding to the outcome domain and subsections corresponding to similar type of initiatives. Our plan for the presentation of results was specified in our protocol and organized in 1/ different sections corresponding to the key concepts detailed in the data-sharing pipeline (intention to data sharing, actual data sharing, results of re-use, output from data sharing, impact of data sharing) and 2/ different subsections corresponding to the different contexts and actors involved in the data sharing pipeline (e.g. targeted group for intention to share data or type of use for re-use of shared data)). A summary of the

data extracted from the included papers was constructed in a tabular form and with basic characteristics and was accompanied by a narrative summary describing all observed results in light of the review objective and *question/s.* Usually, individual studies were summarized in a short text with descriptive statistics of the main results (number, percentages), when appropriate visual representations of the extracted data were provided.

**Patient and public involvement**

There was no patient and public involvement in this scoping review.

**Changes to the initial protocol**

We initially planned to contact leading authors in the field to ask whether they were aware of other unpublished initiatives, but this was not done as it was difficult to identify relevant authors. We found relevant references about data sharing policies including both clinical trials and observational studies, without making a distinction. These references were included in the scoping review and this point was discussed in the text.

**RESULTS**

**Selection of sources of evidence**

A total of 3024 records were identified, 3,005 records (1991 + 1014 in the update) were returned by database searching (2141 without duplicates). An additional 8 records were identified by screening the proceedings of the last three International Congress on Peer Review and Scientific Publication conference abstracts and ten records by snowballing searches. One additional relevant record was identified after screening 630 identified pre-prints. We screened all irrelevant records out by title and abstract, leaving 409 possibly relevant references which were eligible for full-text screening. Subsequently, 316 references were excluded, leaving 93 reports that met the inclusion criteria (**Figure 1**). We inspected websites and contacted when needed 48 initiatives/platforms/journals (we actually screened 49 but SOAR is now integrated into ViVli): 23 data sharing platforms, 13 funding organisation, 5 journals, 5 pre-print repositories and 2 other initiatives. For 33 of these different sources, there was no evaluation component and for 10 additional contacts we received no answer whether they have an evaluation component and/or any data. 4 data sharing platforms (CDSR, YODA, NIDDK, ViVli) and 1 funding organisation (MRC UK) provided some additional data (online metrics and or data about its policy) (**Figure 1**) that were extracted on June 2019 and updated in December 2020.

**Characteristics of source of evidence**

Of the 93 reports, 5 were classified as experimental studies, 58 as surveys, 19 as metrics, 5 as qualitative researches and 6 as others (4 case studies, 1 metrics & survey, 1 metrics and qualitative). The median year of publication was 2018 (range [2001-2020]). The vast majority of these studies were from North America (50, 54%), Europe (16, 17 %) and UK (15, 16%). Eight (9%) were from Asia and 4 (4 %) from Australia. Most (78, 84 %) were focused on data sharing of IPD while the remaining 15 (16 %) adopted a wider definition of material shared (e.g. by including protocols, code). Thirty-eight reports (41 %) were focused on data sharing in publications/journals, 23 (25 %) on data repositories, 8 (9 %) on data sharing by various institutions, 4 (4 %) on trial registries and 20 (21 %) in various other contexts (see **Supplementary Material 3** that presents study characteristics for the detail).

**Collating and summarising the data**

**Figure 2** represents the proportion of the 93 references exploring each outcome domain. In an effort to create a useful synthesis of results, we collated results on each outcomes of each publication and organised them in the pre-specified categories. **Figure 3** presents a detailed overview of the different outcome domains and the related outcomes used in the 93 different references included, organised by type of research.

**Critical appraisal within sources of evidence**

In general, there was a high risk of bias, especially due to study design (e.g. surveys with low response rates and a lack of experimental designs). If available, we tried to present this information in the narrative part of the review.

**Results of individual sources of evidence: intentions to data sharing**

***Clinical Trialists***

Four surveys, investigating intention of data sharing by trialists, reported high data sharing rates of around 75% or higher (see **Figure 4**). These surveys were targeted at authors of published trials and one study at reviewers of a Cochrane group (where the majority of respondents had been involved in a RCT). Studies differed by different estimations of data sharing rates, different selection criteria and/or survey methods. Response rates were comparable between the surveys (42-58%). Reviewers of the Cochrane IPD meta-analysis group were strongly in favour of a central repository and of providing IPD for central storage (83%)[20]. In the survey from Rathi et al.[16], 74% and 72% thought respectively that sharing de-identified data through data repositories should be required and that investigators should be required to share de-identified data in response to individual requests. However, only 18% indicated that they were required by the trial funder to deposit the trial data in a repository. In this survey support for data sharing did not differ by trialist or trial characteristics.[17] Trialists in Western Europe more frequently indicated they have or would share data in order to receive academic benefits or recognition than those from the USA or Canada (58 versus 31%). The most academically productive trialists less frequently indicated they have or would withhold data in order to protect research subjects (24 versus 40% for the less productive), as did those who received industry funding when compared with those who had not (24 versus 43%). The survey from Tannenbaum, 2018[18] suggest that willingness to share may depend on the intended re-use of the data (97% respondents were willing to share data for a meta-analysis versus 73% for a re-analysis). For secondary analyses the willingness to share was largely influenced by respondents' willingness to conduct a similar analysis. In addition, willingness to share was more important after 1 year than after 6 months. In the fourth survey about trials published in Chinese medical journals, the overwhelming majority (87%) stated that they endorsed data sharing.[19]

Intentions to share data for trialists were less important when it comes to data sharing statements in published journal articles (although this section is not specific to clinical trials) (see **Figure 4**). Dependent on the journals considered the rates vary between less than 5 % and until around 25%. An analysis of the first year after the Annals of Internal Medicine policies encouraging data sharing[20] found that data was available without condition for 4%, with conditions for 57%, and unavailable for 38%. Over the first 4 years data was available without condition for 7%, with conditions for 47%, and unavailable for 46% of research articles. Nine percent and 22 % of 160 randomly sampled research articles in the BMJ from 2009 to 2015 made data available or indicated the availability of their data sets.[21] Among 60 randomized cardiovascular interventional trials registered on ClinicalTrials.gov[22] up to 2015 with >5000 enrollment, sponsored by one of the top 20 pharmaceutical companies by 2014 global sales, IPD were available for 15 trials (25%) consisting of 204 452 patients, unavailable for 15 trials (25%) and undetermined for the remaining (50 %) either because of no response or requirements for a full proposal. Reasons for unavailability were: co-sponsor did not agree to make IPD available (4 trials) and trials were not conducted within a specific time (5 trials); for the remaining 6 trials, no specific reason was provided. From 619 RCTs published between 2014 - 2016 in 7 high-ranked anesthesiology journals, only 24 (4%) had a data sharing statement and none provided data in the manuscript or a link to data in a repository.[23] In a survey targeted at the authors of these RCTs, 86 (14%) responded and from 24 participants raw data were obtained. The authors conclude that willingness to share data among anaesthesiology RCTs is very low. From 1 July 2018, clinical trials submitted to ICJME journals must contain a data sharing statement. The reporting of the statement was investigated in a 2 months period before and after this date.[24] The proportion of articles with a data sharing statement was 23% (32/137) before and 25% (38/150) after 1 July 2018, while the number of journals publishing data sharing statements increased from 4/11 to 7/11. Few data sharing statements complied fully with the ICMJE journal criteria, the majority not referring to individual participant data. A total of 300 trials published in 2017-2018 and approximately equal distributed in orthodontics and periodontics were selected, assessed, and analysed with respect to

transparency and reporting.[25] Open data sharing (repository or appendix) was found in 5 % of the trials (11/150 orthodontics and 4/150 peridontics trials). Reproducible research practices and transparency in reproductive endocrinology and infertility (REI) articles was investigated for original articles with a study type mix from REI journals (2013, 2018) and articles published in high-impact general journals between 2013 – 2018.[26] Raw data were available on request or via online database for 1/98 in reproductive endocrinology and infertility RCTs (2013), 0/90 in 2018 and 1/34 in high impact journals. Among a random sample of 151 empirical studies from 300 otolaryngology publications of research studies, using a PubMed search for records published between 1 January 2014 and 31 December 2018, only 5 provided a data availability statement and 3 (2.0%) indicated that data is available.[27]

Intention to share may be even lower when considering data sharing plans of trials registered at clinicaltrials.gov, here the willingness to share data is between 5 and 10%. In a study, 25 551 trial records responded to the Plan to Share IPD field (72%). Of those, 10.9% records indicated « yes » and 25.3% indicated « undecided ».[70] Differences were observed by key funder type, from 11% of NIH funders to 0% of industry answering yes. Importantly, an in-depth review of 154 data sharing plans suggested a possible misunderstanding of IPD sharing with discrepancies found between data sharing plans and declaration of actual data sharing. In a survey, prevalence and quality of IPD sharing statements among 2,040 clinical trials first posted on ClinicalTrials.gov between 01 January 2018 and 06 June 2018 were investigated.[28] The vast majority of trials included in this study did not indicate an intent to share IPD (n = 1,928; 94.5%). Among the trials that did commit to sharing IPD (n = 112, 5.5%), significant variability existed in the content and structure of IPD sharing statements with a need for further clarification, enhanced clarification and better outreach. Data from 287.626 clinical trials registered in ClinicalTrials.gov on 20 December 2018 were analysed with respect to sharing of IPD.[29] Overall, 10.8% of trials with first registration date after December 1, 2015 answered 'Yes' to plan to share de-identified IPD data. The sharing rate ranges von 0% (biliary tract neoplasms) to 72.2% (meningitis, meningococcal) when analysed by disease. For the use case HIV, which was separately analysed, the sharing rate was higher on average (24.5%). In a prediction model, studies that deposit basic summary results in CT.gov, large studies and phase 3 interventional studies are most likely to declare intention to share IPD data.

A 2015 survey,[30] focused on PCORnet (The National Patient-Centered Clinical Research Network), found that a possible barrier toward data sharing intentions related to how data can be used when shared with institutions that have different experience-levels and the possibility of some "competition" between institutions in the marketplace of ideas.

Experimental data suggest that estimations of intent to share data might differ depending on the formulation of the request. For instance, a small randomised prospective study conducted in 2001 including 29 corresponding authors of research publications published in the BMJ, explored their preparedness to share the data from their research.[31] The email contact, randomly allocated was in one of two forms, a general request (asking if the author would `in general' be prepared to release data for re-analysis) and a specific request (a direct request for the data for re-analysis). Researchers receiving specific requests for data was less likely, and slower, to respond than researchers receiving general requests. Similarly, in 2019, a randomized controlled trial in conjunction with a Web-based survey[32] included study authors to explore whether and how much a data-sharing agreement affected primary study authors' attitudes toward their willingness to share IPD. Response rate was relatively low (21 %) in this study since more than 1,200 individuals were initially contacted and 247 responded. Among the responders, study authors who received a data-sharing agreement were more willing to share their data set with an estimated effect size of 0.65 (95% CI [0.39, 0.90]).

Authors of published reports of prevention or treatment trials in stroke were asked to provide data for a systematic review and randomised to be sent either a short email with a protocol of the systematic review attached ('Short') or a longer email that contained detailed information and without the protocol attached ('Long').[33] 88 trials with 76 primary authors were identified in the systematic review, and of these, 36 authors were randomised to Short (trials=45) and 40 to Long (trials=43). Responses were received for 69 trials. There was no evidence of a difference in response rate between trial arms (Short vs Long, OR 1.10, 95% CI 0.36 to 3.33). There was no evidence in response rate and response time between trial arms.

***Trial participants***

Perceptions of trial participants toward data sharing and their intention to share was explored qualitatively. A systematic review with a thematic analysis[34] of 9 qualitative studies from Africa, Asia, and North America identified four key themes emerging from patients: benefits of data sharing (including benefit to participants or immediate community, benefits to the public and benefits to science or research), fears and harms (including fear of exploitation, stigmatization or repercussions, alongside concerns about confidentiality and misuse of data), data sharing processes (mostly consent in the process), and the relationship between participants and research (e.g. trust in different types of research or organization, relationship with the original research team). Some qualitative reports provide data on heterogenous samples including patients and various stakeholders, from low- and middle-income countries. In-depth interviews and focus group discussions involving 48 participants in Vietnam suggested that trials participants may be more willing to be involved in data sharing than trialists.[35] A similar study in a range of relevant stakeholders in Thailand[36] found that data sharing was seen as something positive (e.g. a means to contribute to scientific progress, a better use of resources, a greater accountability, and more outputs) but underlined important reservations including potential harms to research participants, their communities, and the researchers themselves.

In a qualitative study with 16 in depth interviews, cancer patients currently participating in a clinical trial indicated a general willingness to allow re-use of their clinical trial data and/or samples by the original research team, and supported a generally open approach to share data and/or samples with other research teams, but some would like to be informed in this case.[37] Despite divergent opinions about how patients prefer to be engaged, ranging from passive donors up to those explicitly wanting more control, participants expressed positive opinions toward technical solutions that allow indicating their preferences.

Two surveys performed in the US and one in Italy assessed the intention to share rates of trial participants (see **Figure 4**). In one survey[38] with a moderate response rate (47%), 463/799 (58%) of patients favored or strongly favored data sharing while only 9% were against or strongly against it. Most participants (84%) believed that disclosing the data-sharing plan during the informed consent process was important or very important. A higher percentage of minority participants was against data sharing (white, 6%, vs. "other", 13).

In a second survey[39] with a high response rate (79%), 93% were very or somewhat likely to allow their own data to be shared with university scientists and only less than 8% of respondents felt that the potential negative consequences of data sharing outweighed the benefits. Predictors of this outcome were having a low level of trust in people, being concerned about the risk of re-identification or about information theft, having a college degree. 93% and 82 % were very or somewhat likely to allow their data to be shared with academic scientists and scientists in for-profit companies, respectively. Purpose for which the data would be used did not influence willingness to share data except for use in litigation. However, patients were concerned that data sharing might make others less willing to enroll in clinical trials, that data would be used for marketing purposes, or that data could be stolen. Less concern was expressed about discrimination and exploitation of data for profit.

In a survey of Italian patient and citizen groups, 280/2003 contacts provided questionnaires eligible for analysis.[40] 144/280 (51%) had some knowledge about the IPD sharing debate and 60/280 (42%) had an official position. From those who had an official position 35/60 (58%) were in favour and 19/60 (32%) in favour with restrictions. 39% approved broad access by researchers and other professionals and identified information to participants, data de-identification, secure archives, access agreements and sanctions for misuse as important aspects of IPD sharing models.

While consent seems to be a crucial issue for trial participants, an analysis of 98 Informed Consent Forms (ICFs) found that only 6 (4%) indicated a commitment to share de-identified IPD with third party researchers.[41] Commitments to share were more common in publicly funded trials than industry-funded trials (7% vs 3%).

***Publishers/funders***

*Publishers*

Several studies were found about intentions (and data sharing policies) of publishers. Many publishers have developed data sharing policies (20-75%), however, less than 10% are mandatory (see **Figure 4**). In a 2009 survey[42] of editors of different member journals of the World Association of Medical Editors (WAME) (response rate 22%), 2% and 19% of journals required participant level data and authors to specify their data-sharing plan, respectively. A similar survey of 10 high-impact surgical journals during 2009 and 2012 found only one

journal that had a mandating data sharing policy.[43] Data sharing statements were only found in 2/246 (1%) RCTs published in these 10 journals. Another study from a random sample of 60 journals[44] found that 21 (35 %) provided instructions for patient level but only 4 (7 %) mandated sharing IPD (all were oncology journals). A review of 88 websites from dental journals[45] suggested that 17 accepted raw data as a complementary material. A 6-year cross-sectional investigation of the rates and methods of data sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018 was performed.[46] 8/14 (57.1%) journals had data sharing policies for published RCTs. From the included 394 RCTs zero shared their data publicly. 40/60 clinical psychology journals had a specific policy for data sharing (2017).[47] Only one journal mandated data sharing, while 37 recommended it. The findings suggest great heterogeneity in journal policies and scarce enforcement. Online instructions for authors from 38 high impact addiction journals were reviewed for 6 publication procedures, including data sharing (2018). 28/38 (74%) of the addiction journals had a data sharing policy, none was mandatory.[48] It was concluded that many addiction journals have adopted publication policies but more stringent requirements have not been widely adopted. Instructions for authors of 43 high impact nutrition and dietetics journals were reviewed with respect to procedures to increase research transparency (2017).[49] 25/33 (75%) journals publishing original research and 4/10 review journals had a data sharing policy Among 109 peer-reviewed and original research-oriented dental journals that were indexed in the MEDLINE and/or SCIE database in 2018, a data sharing policy was present in 32/109 (29.4%) and 2 of those had a mandatory policy.[50] It is concluded by the authors that currently data sharing policies are not widely endorsed by dental journals. In a cross-sectional survey 14 ICJME-member journals and 489 ICJME-affiliated journals that published a RCT in 2018 were evaluated with respect to data sharing recommendations.[51] 8/14 (57%) of member journals and 145/489 (30%) of affiliated journals had an explicit data-sharing policy on their website. In RCTs published in member journals with a data sharing policy, there were data-sharing statements in 98/100 (98 %) with expressed intention to share individual patient data reaching in 77/100 (77%). In RCTs published in affiliated journals with an explicit data-sharing policy, data-sharing statements were rare 25/100 (25%), and expressed intentions to share individual participant data were found in 22/100 (22%).

Changes in policies from 2013 to 2016 regarding public availability of published research data were investigated for 115 paediatric journals.[52] In 2012 77 /115 (67%) and in 2016 56/115 (49%) accepted storage in thematic or institutional repositories. Publication of data on a website was accepted by 27/115 (23%) and 15/115 (13%). Most paediatric journals recommend that authors deposit their data in a repository but they do not provide clear instructions for doing so.

*Funders and clinical trial units*

Several studies investigated mandatory data sharing policies of funders. 30-80% of the non-commercial funders provide data sharing policies, the highest rates were observed in the US. Only around 10-20% of these policies are mandatory (see **Figure 4**). In one study 50% of the top non-commercial funders had a data sharing policy but it was found that only in 2/20 cases data sharing is required. Six funders offered technical or financial resources to support IPD sharing.[53] Trial transparency policies were investigated for 9/10 top non-commercial funders in the US (May to November 2018).[54] 7/9 (78%) funders had a policy for individual patient data sharing, for 1 it was mandatory. 6 offered data sharing and 5 monitored compliance. From 96 responders of 190 contacted non-commercial funders in France, 31 were identified to fund clinical trials (2019).[55] 9/31 (29%) had implemented a data sharing policy. Among these 9 funders, only one had a mandatory sharing policy and 8 a policy supporting but not enforcing data sharing. Funders with a data sharing policy were small funders in term of total financial volume.

Three studies investigated mandatory data sharing policies for commercial sponsors (see **Figure 4**). In a 2016 survey, 22/23 (96%) companies among the top 25 companies by revenue had a policy to share IPD[59]. Of a second sample of 42 unselected companies, 30 (71 %) had one. These policies generally did not cover unlicensed products or trials for an off-label use of a license product. 52 % of top companies, and 38 of the sample including all companies considered requests for IPD on additional trials not explicitly covered by their policy.[56] A second survey[57] studied data availability for 56 publications reporting on 61 industry-sponsored clinical trials of medicines. Of those 61 studies, 32 (52%) had a public data sharing policy/process.

78 non-commercial and a random sample of 100 commercial funders (selected from top 100 pharmaceutical companies in terms of drug sales) of clinical research having funded at least one RCT in the years 2016 to 2018 were surveyed (15 February 2019 – 10 September 2019).[58] 30/78 (38%) non-commercial funders had a data-

sharing policy with 18/30 (60%) making data-sharing mandatory and 12/30 (40%) encouraging data-sharing. 41/100 (41%) of commercial funders had a data-sharing policy. Among funders with a data-sharing policy, a survey of two random samples of 100 RCTs registered on Clinicaltrial.gov, data-sharing statements were present for 77/100 (77%) and 81/100 (81%) of RCTs funded by non-commercial and commercial funders respectively. Intention to share data was expressed in 12/100 (12%) and 59/100 (59%) of RCTs funded by non-commercial and commercial funders. The survey indicated suboptimal performance of funders in setting up data sharing policies.

Among 23 UK Clinical Research Collaboration (UKCRC) registered Clinical Trial Units (CTUs)[10] (response rate = 51 %), 5 (22 %) had an established data sharing policy and 8 (35%) specifically requested consent to use patient data beyond the scope of the original trial (see table). Concerns were raised about patient identification, misuse of data, and financial burden. No CTUs supported the use of an open access model for data sharing.

A 2005 survey[59] over 107/122 accredited medical schools in the United States (response rate = 88%) explored data sharing in the context of contractual provisions that could restrict investigators' control over data in the context of industry funded trials. There was a poor consensus among senior administrators in the offices of sponsored research at these institutions when it turned to prohibiting investigators from sharing data with third parties after the trial is over (41 % allowed it, 34 % disallowed it, and 24 % were not sure whether they should allow it).

In a survey targeted at European heads of imaging departments and speakers at the clinical trials in radiology sessions (July – September 2018), response rate was 132/460 (29%).[60] Responses were received from institutions in 29 countries, reporting 429 clinical trials. For future trials, 98% of respondents (93/95) said they would be interested in sharing data, although only 34% had shared data already (23/68). The major barriers to data sharing were data protection, ethical issues, and lack of a data sharing platform.

**Results of individual sources of evidence: actual data sharing**

***Re-users***

*Studies related to journal articles*

Several studies have been performed investigating data sharing rates for studies that have been published in journals, the majority with data sharing policies and high impact (**Figure 5**). Even with strict data sharing policies, the data sharing rates are low or at maximum moderate and vary between 10 and 46%, except for one study with a very high data sharing rate due to a partly preselected sample of authors willing to share their data[18] . In the 6-year cross-sectional investigation of the rates and methods of data sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018, none of 394 included clinical trials shared their data publicly.[46] From 86 responders in a survey targeted at the corresponding authors of 619 RCTs published between 2014 - 2016 in 7 high-ranked anesthesiology journals, raw data were only obtained for 24 studies.[23] 62 declined to share raw data. In a study targeted at PLOS Medicine and PLOS Clinical Trials publications, performed in 2009, 1/10 (10%) of data sets were made available after request[28]. In articles in Chinese and international journals from 2016, a sharing behaviour was indicated for 29/247 (11%) of articles.[19] From top 10 general and internal medical journals investigated in 2016, IPD was provided after request for 9/61 (15%) of pharma sponsored studies [56]. For BMJ research articles, published between 2009 and 2015, data sets were made available in 7/157 (4%) of the articles.[30] For the sub-sample of clinical trials the rate was higher (5/21 (24%)). From 317 clinical trials published in 6 general medical journals between 2011 and 2012, 115 (36%) granted access to data[35] . The data availability for RCTs published in BMJ and PLOS Medicine between 2013 and 2016 was 17/37 (46%)[42].

In a parallel group RCT an intervention group (offer for an Open Data Badge for data sharing) was compared with a control group (no badge for data sharing).[61] The primary outcome was the data sharing rate. From 160 research articles published in BMJ Open, 80 were randomised to the intervention and control group, of which 57 could be analysed in the intervention and 54 in the control group. In the intervention group data was available at a third-party repository for 2/57 (3.5%) and upon request for 32/57 (56.1%), respectively in the control group: 3/54 (5.6%) and 30/54 (56%). Data sharing rates were low in both groups and not different between the groups.

*Data sharing for IPD meta-analyses*

Some examples demonstrate that data availability for IPD meta-analyses is still limited despite the various data sharing initiatives/platforms (**Figure 5**). The availability can be increased under specific circumstances, such as building up a disease-specific repository for a scientific community, as demonstrated for a repository of IPD from multiple low back pain RCTs with IPD from 20/42 (48%) RCTs included[57] and studies on anti-epileptic drugs collected by a Cochrane group with IPD of 15/39 (38%) studies included[40]. In another study, from different databases 35 individual participant data meta-analyses with more than 10 eligible RCTs were identified (May 1, 2015 to February 13, 2017).[61] From 774 eligible RCTs identified in these meta-analyses, 517 (66.8 %) contributed data. The country where RCTs are conducted (UK versus US), impact factor of the journal (high versus low) and recent RCT publication year were associated with higher sharing rates. In three other studies, the availability of datasets for IPD meta-analysis was limited (0-17%). In one study, dedicated to one commercial sponsor with one specific medicinal product performed in 2014, IPD from 24 trials were requested without success[47] From 15 requests (13 direct to authors, 2 to a repository) in 2014/2016, IPD was received for 2/15 (13%) of the studies[51]. From 217 RCTs published since 2000 in orthopedic surgery, agreement to send IPD was achieved in 37/217 (17%)[35].

The low data availability for IPD-meta-analyses is underlined by two experimental studies. One experimental study covered the issue of actual data sharing. In the small randomized prospective study,[31] where 29 corresponding authors of original research articles in a medical journal were contacted via two different modes (general versus specific request), only one author actually sent the data immediately in response to a specific request and one author, without caveats, reported a preparedness to send the data in response to a general request.

A randomized controlled trial investigated the effect of financial incentives on IPD sharing.[62] All study participants (129 in total) were asked to provide the IPD from their RCT. Those allocated to the intervention group received financial incentives, those from the control group not. Primary outcome was be the proportion of authors who provide IPD. None of the authors shared their IPD, whatever the group.

Two studies investigated the completeness of data availability in IPD meta-analyses. Out of 30 IPD meta-analyses included in a survey,[63] 16 did not have all the IPD data requested. The assess rate of retrieving IPD for use in IPD-meta-analyses was investigated in a systematic review.[64] Only 188 (25%) of 760 IPD meta-analyses retrieved 100% for of the eligible IPDs for analysis and there was insufficient evidence that IPD retrieval rates improved over time.

*Access to repositories/platforms*

Only a few studies describe access to repositories/platforms from the viewpoint of the user (**Figure 5**). Experiences with two major platforms (CSDR, PDS) were reported.[65] In these very early phase of the projects, no data access was possible with CSDR, faster data acquisition was achieved via the Project Data Sphere. High sharing rates were reported from academic repositories (MRC CTU, BioLINCC), From 103 requests to MRC CTUs, access was granted in 80/103 (78%) cases[22] . In a survey of investigators 536/536 (100%) received access to BioLINCC during a time period between 2007 and 2014[31].

***Repositories/platforms***

*Commercial sponsors*

Different initiatives and platforms have been initially implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now opened to academic trials but this has not been used quite often so far). This covers the YODA project, CSDR, ViVli and SOAR (which is now part of ViVli). For the different platforms and repositories metrics describing the actual use of the data are available (**Figure 5**).

6 studies have accessed data sharing rates for CSDR. From 2014 to the end of January 2019, there were a total of 473 research proposals submitted to CSDR.[66] Of these, 364 met initial administrative and data availability checks, and the independent review panel approved 291. 222/473 (46.9%) of the requests gained access to the data (in progress and completed). Of the 90 research teams that had completed their analyses by January 2018, 41 reported at least one resulting publication to CSDR. Less than half of the studies ever listed on CSDR have been requested. Between 2014 and 2017 CSDR received a total of 172 research proposals, of which 105 (61%) were approved[26]. In another study focussing on availability and use of shared data from cardiometabolic clinical trials in CSDR, covering the time period between 2013 and 2017, 198 (62%) were approved with and without conditions[18]. In year one of the use of CSDR (2013-2014), 36 research proposals were approved with conditions, of these 23 (64%) had progressed to a signed data-sharing agreement[24]. During 2014 through 2017, Boehringer-Ingelheim listed 350 trials for potential data sharing at CSDR.[67] 55 research proposals were submitted, of which 37 (67.3%) were approved. All approved research proposals submitted to Boehringer-Ingelheim, except one, addressed new scientific questions or were structured to generate new hypotheses for further confirmatory research, rather than replicating analyses by the sponsor to affirm previous research. Between 2013 and 2015 177 research proposals were submitted to CSDR, of which access was granted for 144 (81%) of these proposals[23].

In the first year since launch October 2014, YODA received 29 requests of which all were approved (100%) [49], Experience with the YODA project in 2017 reported 73 proposals of which 65 had been approved (Ross, 2017). A more recent publication reports the metrics for data sharing of Johnson & Johnson clinical trials in the YODA project until August 27, 2018.[68] 100 data requests have been received from 89 principal investigators for a median of 3 trials per request. 90/100 requests (90 %) have been approved and a data use agreement has been signed in 82/100 (82%).

The use of the open access platforms CSDR, ODA and SOAR together between 2013 and 2015 was investigated in a study. Of the 234 proposals submitted, 154 (66%) proposals were approved[41]

The data available show that the use of these platforms has increased steadily since its initiation and that 50% and more of the data requests lead to actual data sharing. The reasons for not sharing are manifold but rarely data access is denied by the platforms. Our assessment of CDSR, YODA, NIDDK and ViVli websites is presented in **Table 1**.

| Platform | Metrics date | Available studies | Number of requests | Number of requests with data shared | Number of requests with data leading to publication | Number of publications |
|---|---|---|---|---|---|---|
| CSDR | 30/11/2020 | 3008 | 621 | 318 | 59* | 79 |
| YODA | 15/11/2019 | 334 | 196 | 173 | 29 | 35 |
| ViVli | 02/11/2020 | 5203 | 215 | 123 | 8 | 9 |

**Table 1**: Metrics of CDSR, YODA, and ViVli websites
NIDDK had also some metrics concerning the number of request (530) but no other information
*publication anticipated

Ethics approval in applications for open-access clinical trial data from CSDR was investigated in a survey.[69] Projects with and without ethics approval were approved at roughly similar rates (62/111 and 43/61).
The proportion of trials, where Pharma and medical device industry provide IPD for secondary analyses and thus the completeness of trial data is still limited.[57] Only 15% of 61 industry-sponsored clinical trials were available 2 years after publication. For companies listing at least 100 studies at CSDR a search was performed in ClinicalTrials. gov (1/2016, studies terminated/ completed at least 18 mo. before search date).[70] From 966 RCTs registered in ClinicalTrials.gov, only 512 (53%) were available in CSDR and only 385 (40%) of the RCTs were registered and listed at CSDR with all datasets and documents available. This was the case despite delay of 18 mo. since the completion of drug trials by the company sponsor. Differences between sponsors were observed. Pharma repositories may cover only part of the trials with commercial sponsors needed for meta-analyses. In a study, investigating data availability for industry-sponsored cardiovascular RCTs with more than 5000 patients,

performed by a top 20 pharma company and registered at ClinicalTrials.gov (up to Jan. 2015), for only 25% of the trials identified data was confirmed to be available.[22] In 50% of the cases availability could not be definitely confirmed.
As part of the Good Pharma Scorecard project, data sharing practices were assessed for large pharmaceutical companies with novel drugs approved by the FDA in 2015, using data from CT.gov, Drugs@FDA, corporate websites, data sharing platforms and registries (e.g. YODA, CSDR).[71] 628 trials were analysed. 25% of large pharma companies made IPD accessible to external investigators for new drug approvals, this proportion improved to 33% after applying a ranking tool.

*Non-commercial sponsors*

Disease-specific academic clinical trial networks have a long history of IPD sharing, especially with respect to US related NIH institutions. This is clearly demonstrated by the available literature; however, the metrics of data sharing is not always as transparent as by the industry platforms and cannot be structured and documented easily in a table.
In a survey on the use of the National Heart, Lung, and Blood institute Data Repository, access to 100 studies initiated between 1972 and 2010 was investigated.[72] A total of 88 trial datasets were requested at least once and the median time from repository availability and the first request was 235 days.
Since its inception in 2006 and through October 2012, nearly 1700 downloads from 27 clinical trials have been accessed from the Data Share website belonging to the National Drug Abuse Treatment Clinical Trial Network (CTN) in the United States, with the use increasing over the years.[73] Individuals from 31 countries have downloaded data so far.
In a case study approach, the data sharing platform Data Share from the National Institute of Drug Abuse (NIDA) was investigated in detail.[74] As of March 2017, the Data Share platform included 51 studies from two trial networks (36 studies from CTN and 15 studies from NID Division of Therapeutics and Medical Consequences). From 2006 through March 2017, there have been 5663 downloads from the Data Share website. Of those, 4111 downloads have been from the US.
The Project Data Sphere (PDS) is an open source data sharing model that was launched in 2014 as an independent, non-profit initiative of the CEO roundtable on cancer.[75] PDS contains data from 72 oncology trials, donated by academic, government, and industry sponsors. More than 1400 unique researchers have accessed the PDS database more than 6500 times. As an example, a challenge to create a better prognostic model for advanced prostate cancer was issued in 2014, with 549 registrants from 58 teams and 21 countries.
The Immune Tolerance Network (ITN) is a National Institute of Allergy and Infectious Diseases /National Institutes of Health-sponsored academic clinical trial network.[76] The Trial share portal, which was released for public access in 2013, provides complete open access to clinical trial data and laboratory studies from ITN trials at the time of the primary study publication. Currently, data from 20 clinical trials are available and additional 17 will be released to the public at the timepoint of first publication. So far, more than 1000 downloads have been registered.
In the MRC Clinical Trials Transparency Review Final Report (November 2017), the MRC UK reported that 24/107 (22%) trials which started during the review period had created a database or collection for sharing. Seven of these datasets (7/24, 29%) had already been shared with other researchers.[77]
From 215 requests submitted for PLCO data, 199 (93%) were approved, for NLST 214 (89%) from 240 requests.[78]

*Other stakeholders*

In a case study about experiences with data sharing among data monitoring committees, access to five concurrent trials assessing the level of arterial oxygen, which should be targeted in the care of very premature neonates, was investigated.[79] The target of taking account of all relevant evidence when monitoring the clinical trials, could only partially be reached.
One case-study addressed directly the issue of costs. Data from two UK publicly funded trials were used to assess resource implications of preparing IPD from a clinical trial to share with external researchers.[80] One trial, published in 2007, required 50 hours of staff time with a total estimated cost of £3185, the other published in 2012 required 39.5 hours with £2540.

**Results of individual sources of evidence: re-use**

***Any type of re-use***

The majority of research projects using shared clinical trial data are dealing with new research. This covers studies of risk factors and biomarkers, methodologic studies, optimizing treatment and patient stratification studies and subgroup analyses. So far only some IPD-meta-analyses have been planned and a few reported. Re-analyses are only exceptionally applied.

Early experience at CDSR, involving Glaxo Smith Kline trials[81] found low rates of IPD meta-analysis and re-analyses, the vast majority being secondary analyses (studies of risk factors or biomarkers, methodologic studies, predictive toxicology or risk model, studies of optimizing treatments, subgroup analyses etc.). Similar results were found in an update of the analysis.[82]

In the YODA project, which received 73 proposals for data sharing as of June 2017 and approved 65 proposals,[83] the most common study purposes were to address secondary research questions (n=39), combine data as part of larger meta-analyses (n=35) and/or validate previously published studies (n=17).

Among the 172 requests to the National Heart, Lung and Blood Institute (NHLBI) data repository with online project descriptions and a coded purpose, 72% of requests were initiated to address a new question or hypothesis, 7% to perform a meta-analysis or combined study analysis, 2% to test statistical methods, 9% to investigate methods relevant to clinical trials, and 9% for other reasons.[72] In only two requests, the available description suggested a re-analysis.

From 2014 to the end of January 2019, 222/473 (46.9%) of the requests to CSDR gained access to the data (in progress and completed).[66] 90/222 (40.5 %) research teams had completed their analyses by January 2018. 41 have published at least one paper, with another 28 that were expected to publish soon.

In the SPRINT challenge. Individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and identify novel scientific or clinical findings.[84] Among 200 qualifying teams, 143 entries were received.

***Further additional analyses***

There were few indications concerning the exact type of secondary analysis that was performed. Approved proposals by subject area are available for the Cancer Data Access system (CDAS), covering two large cancer screening trials (Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial and National Lung Screening Trial (NLST).[78] From the 199 approved requests to PLCO between November 2012 and October 2016, 84 (42%) were dedicated to cancer etiology, 66 (33%) to trial-related screening, 29 (15%) to other object areas, 14 (7%) to risk prediction and 6 (3%) to image analysis. From the 214 approved requests to NLST, 95 (44%) were dedicated to image analysis, 90 (42%) to trial-related screening, 14 (7%) to other subject areas, 10 (5%) to cancer etiology and 5 (2%) to risk prediction.

***IPD meta-analyses***

In a study, IPD meta-analyses seems to represent a small proportion of data re-use. Among the 174 research proposals approved of up to 31 August 2017 by CSDR, 12 proposals were IPD meta-analysis, including network meta-analysis.[85] All were retrospective IPD meta-analyses (i.e. none was prospective IPD meta-analysis).

***Re-analyses***

A 2014 survey of published re-analyses[86] found that a small number of reanalyses of RCTs have been published (only 37 re-analyses of 36 initial RCT) and only a few were conducted by entirely independent authors. Thirty-five percent of these reanalyses led to changes in findings that implied conclusions different from those of the original article about the types and number of patients who should be treated.

In the survey of 37 RCTs from The BMJ and PLOS Medicine[87] published between 2013 and 2016, 14 out of 17 (82%, 95% IC: 59% to 94%) available studies were fully reproduced on all their primary outcomes. Of the remaining RCTs, errors were identified in two but reached similar conclusions and one paper did not provide enough information in the Methods section to reproduce the analyses.

**Results of individual sources of evidence: output from data sharing**

Publications can be considered at the main research output of data sharing. Publication activity of re-use of clinical trial data has been considered in several studies. Detailed data are available for academic clinical trial networks and disease-specific repositories in the US, some of them practising data sharing already for a period longer than 10 years. Here, up to moderate publication output has been observed dependent on the individual repository. So far this is not the case for the repositories storing clinical trial data from commercial sponsors, taking into consideration that these repositories were established around five years ago and that there is usually a considerable time lag between request, approval, analysis and publication. Current statistics are indicating improvement of publication output with time.

***Non-commercial sponsors***

In a cross-sectional web-based survey about access to clinical research data from BioLINCC, covering the period from 2007 to 2014, 98 out of 195 responders (50%) reported that their projects have been completed of which 66 (67%) have been published.[88] Of the 97 respondents who have not yet completed their proposed projects, 81 (84%) explained that they plan to complete their project; 63 (65%) indicated that their project was in an analysis/manuscript draft phase.

In a survey targeted at European heads of imaging departments and speakers at the clinical trials in radiology sessions (July – September 2018), 23/68 reported that they had shared data already.[60] At least 44 original works were published based on the data shared by the involved 23 institutions.

In five studies (**Table 2**) the number of publications has been reported, usually referring to the number of trials included in the repository/platform.

| Reference | Repository/ platform | No. of trials included in repository/platform | No. of published articles | Assessment |
|---|---|---|---|---|
| Shmueli-Blumberg, 2013 | CTN Data Share | 27 trials (1700 downloads) | 13 | 2012 |
| Zhu, 2017 | CDAS | 2 trials (PLCO, NLST) (455 requests) | 25% for PLCO projects, 19% for NLST projects | 2016 |
| Coady, 2017 | BioLINCC | 100 trials (88 requested at least once) | 35% of clinical trials at least 1 publication 5 years after availability in the repository | 5/2016 |
| Huser, 2018 | NIDA Data Store | 51 trials | 14 | 3/2017 |
| Pisani, 2017 | WWARN | 186 trials | 18 | 2016 |

**Table 2**: Studies reporting published outputs for non-commercial sponsors

***Commercial sponsors***

Various studies explored Metrics of both YODA and CSDR (**Supplementary Material 4**).

Up to 2021, ViVli's website indicates very few published output. We were not able to retrieve published output from NIDDK. **Figure 6** presents publication metrics from CSDR (up to 31 August 2019) and YODA (up to 1st July 2019). Over 88 published papers (62 from CSDR and 26 from YODA), 49 were secondary analyses (42 from

CSDR and 7 from YODA), 30 were meta-analyses (13 from CSDR and 17 from YODA), 6 were methodological studies (5 from CSDR and 1 from YODA) and 3 were re-analyses (2 from CSDR and 1 from YODA). The detail of these publications[82 83 89] is presented in **Supplementary Material 5**.

**Results of individual sources of evidence: impact of research output**

Evidence on the impact of research output from sharing IPD of clinical trials is still very low. So far only two studies, with inconsistent results, could be identified, dealing with this issue and focusing only on citation metrics.

One study, already published in 2007, suggested that sharing detailed research data was associated with an increased citation rate.[90] From 85 cancer microarray clinical trials, published between January 1999 and April 2003, 41 made their microarray data publicly available on the internet. For 2004 – 2005, the trials with publicly available data received 85% of the aggregate citations. Publicly available data was significantly associated with a 69% increase in citation, independently of journal impact factor, date of publication and author country of origin.

Citation metrics from 224 publications based on repository data of clinical trials from the NHLBI Data Repository were compared with publications that used repository observational study data as well as a 10%-random sample of all NHLBI-supported articles published during the same period (January 2000 – May 2015).[72] Half of the publications based upon clinical trial data had cumulative citations that rank in the top 34% normalized for subject category and year of publication compared to 28.3% for the publications based on observational studies and 29% for the random sample. The differences were, however, not statistically different.

In the SPRINT challenge. Individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and identify novel scientific or clinical findings.[84] Among 200 qualifying teams, 143 entries were received. Entries were judged by a panel of experts on the basis of utility of the findings to clinical medicine, originality and novelty of the findings, and quality and clarity of the methods used. All submissions were also open for crowd voting among the 16,000 persons following the SPRINT Challenge. Cash prizes were awarded, and winners were invited to present their results. 143 entries to the SPRINT data challenge were received.

**DISCUSSION**

**Summary of evidence**

There are major differences with respect to the intention to share IPD from clinical trials between the different stakeholder groups. The studies available so far show that clinical trialists and a bit less expressed study participants, as one the main actors of clinical trials, usually have a high willingness to share data (60-80%). This is much less developed when it comes to data sharing statements published in journal articles. Dependent on the journals considered, the rates vary between less than 5% until around 25%. This is even worse when data sharing plans documented in registries (e.g. CT.gov) are analysed. Here the willingness to share data is between 5 and 10%.

As a consequence, a large discrepancy between the positive attitude towards data sharing in general and the intention to do so in a concrete study has to be ascertained. Publishers, enabling the publication of research output from clinical trials and funders/sponsors, financing clinical trials, could be major drivers to change the situation. Meanwhile many publishers have developed data sharing policies (20-75%), however, less than 10% are mandatory and have thus not been enforced. There are differences between the journals with some of the high impact journals being stronger than the others *(e.g. JAMA, NEJM, PLOS Medicine, BMJ).* For funders, the situation is similar but different between commercial and non-commercial funders. 30-80% of the non-commercial funders provide data sharing policies with US and NIH at the front. Only around 10 to 20% of these policies are mandatory. Data sharing policies have been developed more often in the group of commercial funders (40-95%) but information on the proportion of mandatory policies is missing. In summary, the pressure by publishers and funders to share data is still limited and the situation is only slowly improving. The situation

is better for the pharmaceutical industry, which has not only promoted data sharing policies in their organisations to a large degree but has also implemented platforms and repositories, practically supporting the process of data sharing (e.g. CSDR, Yoda, vivli).

Several studies have been performed investigating data sharing rates for clinical studies that have been published in journals. A focus has been on high impact journals with strict data sharing policies (e.g. PLOS Medicine, BMJ, Ann Intern Med), demonstrating data sharing rates between 10% and 46%, except for one study with a very high data sharing rate due to a partly preselected sample of authors willing to share their data. Data availability for IPD meta-analyses is usually limited (0-20%), only under specific circumstances (Cochrane group, disease-specific repository) the availability can be increased to 50% and more. A few individual studies describe access to repositories/platforms from the viewpoint of the user, not allowing identification of a general pattern. Different initiatives and platforms have been implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now opened to academic trials but this has not been used quite often so far). This covers the YODA project, CSDR, ViVli and SOAR (which is now part of ViVli). The data available show that the use of these platforms has increased steadily since its initiation and that 50% and more of the data requests lead to actual data sharing. The reasons for not sharing are manifold but rarely data access is denied by the platforms.

The majority of research projects using shared clinical trial data are dealing with new research. This covers studies of risk factors and biomarkers, methodologic studies, optimizing treatment and patient stratification studies and subgroup analyses. This is important because new research may be easier to publish in peer-reviewed journals, which is a major driver of academic career.

So far only some IPD meta-analyses have been planned as part of data sharing initiatives and a few reported. There are many hurdles for IPD meta-analyses, including the findability, the accessibility and the re-usability of datasets (F, A and R in FAIR). ECRIN has developed a metadata dictionary (MDR), able to identify clinical studies and data objects related to it (e.g., protocol, DMP, CRF).[91] With this tool studies that can be discovered for which datasets are available and what the conditions for access are (ECRIN, MDR). Even if IPD datasets are accessible for meta-analyses, the studies are usually distributed over various repositories. This has been demonstrated in several studies of our scoping review. One central repository could simply the situation but instead, the number of repositories is steadily increasing.[2] The situation could be considerably improved with more standardisation and harmonisation of data and procedures and a federating approach between repositories.

Re-analysis of clinical trials data may help the scientific community to access the validity of reported trial results. An illustrative example is the restoring study 329, investigating efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. The re-analysis came to different conclusions with important implications for both clinical practice and research.[3] RIAT (Restoring invisible & abandoned trials support center) was initiated as an international effort to tackle bias in the way research is reported with the goal of providing more accurate information to patients and other healthcare decision makers.[92]

One of the problems that is tackled by RIAT is misreporting (inaccurately or incompletely reported trials). In our scoping review we found that re-analyses are only exceptionally applied. In one review, the majority of studies was reproduced on all primary outcomes, in another around one third of studies led to changes in findings different from the original articles. It seems to be that re-analysis is only attractive in a minority of cases deserving major public interest. Nevertheless, for those cases repositories holding and sharing IPD could be of major help and speed up the process of data sharing. It could be of interest to make a link between RIAT and data sharing platforms and initiatives.

Publications can be considered as the main output from data sharing. Usually, there is a considerable time lag between requesting data for re-use, receiving shared data, performing secondary analysis, writing a manuscript and publishing the secondary analysis. This has to be taken into consideration when the publication output of data sharing initiatives and platforms is analysed. Repositories and platforms mainly dedicated to commercial trials now exist for around 5 years, so only a limited publication output can be expected. Fortunately, these repositories provide detailed metrics for the data sharing requests, including number and type of publications originating from data sharing. As expected, the number of publications related to data sharing for commercial studies is still limited, however, current statistics indicate improvement over time. The situation with non-commercial sponsors is different. Some academic clinical trial networks and disease-repositories have been

successfully implemented (mainly in the US) and practice data sharing already for quite a long time, some more than 10 years. Here data sharing is part of the research culture and the exchange of data is based upon elements such as trust, technical support and common benefit. Outstanding examples are BioLINCC,[88] NIDA[73] and WWARN.[93] This is reflected in the data sharing rates for IPD meta-analyses, which are rather low if data requests are targeted at authors directly compared to data sharing requests within communities (e.g. Cochrane groups) or related to specific repositories. Outside clinical trial networks and disease-specific repositories, data sharing of IPD is still very limited. Possible reasons are the lack of widely accepted repositories for non-commercial clinical trials and insufficient incentives and benefits related to data sharing.
One issue is that not for all projects the publications from secondary analysis are regularly updated, so statistics may be biased. Improvements could be achieved with a prospective registration of any protocol of secondary data use similar to the trial registries (e.g. CT.gov), a mandatory link between the registration and the original publication or data set and the necessity to refer to the primary publication or dataset if the re-analysis is published. Existing approaches and tools could then be extended to automatically identify publications related to re-use of data and make a link to the original work (e.g. see crossmark – crossref[94] metadata repository (MDR) developed by ECRIN linking clinical studies with related data objects).[91] Another possibility could be to set up a register for secondary analyses.

To be widely accepted, research output from shared data should have an impact on medical research (e.g. generation of new hypotheses) and medical health (e.g. changing treatment via guidelines). It is well known that the impact of primary studies on medical research and health has often considerable time-lag and direct effects are not easy to demonstrate. So, it is to be expected that the proof of evidence of research output from shared data is even more difficult to demonstrate. In this scoping review, taking into consideration the limited time available for data sharing activities to generate impact, no major effects could have been expected. As a consequence, the evidence on impact of data sharing is still very low. This may mean that it is still too early to measure any impact or that the impact is very limited. So far, only surrogate measures have been considered (citation metrics) with inconclusive results. It is hoped that in the next years, more studies with more relevant criteria and metrics are performed. One option could be to closely follow-up the SPRINT challenge, where 143 secondary analyses on one clinical trial were performed and it would be nice to see whether one or more of these secondary analyses really had an impact.

**Limitations**

Retrieving and synthetizing information for this study proved to be difficult because we operated in a very siloed landscape where each initiative platforms operates with its own metrics. We tried to be exhaustive by reviewing both the literature and the most important initiatives. However, it is hard to keep the review up-to date as we are studying a moving target in a rapidly changing environment with more and more new initiatives. In addition, data sharing has not a long history and many of the initiatives and activities have been launched in the near past. Therefore, only a limited research output from data sharing can be expected so far and indeed, the number of publications is disappointing. It is expected and can already be seen that the number of publications will increase.

**Conclusions**

There is currently a gap in the evidence base evaluating impact of IPD sharing, which causes uncertainties in the implementation and adoption of current data-sharing policies. Data sharing faces many challenges including, for instance, the scepticism of trialists.[95] There is therefore a need to provide high level evidence that the value of medical research liable to inform clinical practice increases with greater transparency and the opportunity for external researchers to re-analyse, synthesize, or build on previous data. First, a register (such as PROSPERO[96]) for any secondary use of shared data has to be created. The inscription in such a register could be mandatory for any data sharing agreement/publication such as registration of clinical trials. Such a register would make it possible to build easily an observatory of data sharing practices providing direct feedback without the actual silos we had to face. In addition, such a register may help to prevent any selective publication of secondary analyses. Lastly, we suggest that interventional studies have to be run to determine the optimal data sharing policy and/or incentives that adds value for clinical research. It has, however. to be taken into consideration that the experimental studies performed so far were not very conclusive, indicating that experimental studies in this area are very demanding.

**REFERENCES**


1. Nosek BA, Alter G, Banks GC, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. *Science (New York, NY)* 2015;348(6242):1422-5. doi: 10.1126/science.aab2374 [published Online First: 2015/06/27]
2. Banzi R, Canham S, Kuchinke W, et al. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials* 2019;20(1):169. doi: 10.1186/s13063-019-3253-3 [published Online First: 2019/03/17]
3. Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. [published Online First: 2015/09/18]
4. Data sharing and the future of science. *Nature Communications* 2018;9(1):2817. doi: 10.1038/s41467-018-05227-z
5. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data: A Proposal from the International Committee of Medical Journal Editors. *PLoS Med* 2016;13(1):e1001950. doi: 10.1371/journal.pmed.1001950 [doi]
PMEDICINE-D-15-03676 [pii] [published Online First: 2016/01/21]
6. Rosenbaum L. Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp. *N Engl J Med* 2017;376(23):2201-03. doi: 10.1056/NEJMp1704482 [doi] [published Online First: 2017/04/27]
7. Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *PLoS Med* 2017;14(6):e1002315. doi: 10.1371/journal.pmed.1002315 [doi]
PMEDICINE-D-17-01550 [pii] [published Online First: 2017/06/06]
8. Vasilevsky NA, Minnier J, Haendel MA, et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* 2017;5:e3208. doi: 10.7717/peerj.3208 [published Online First: 2017/05/04]
9. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7(12):e018647. doi: 10.1136/bmjopen-2017-018647 [published Online First: 2017/12/17]
10. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. *Journal of clinical epidemiology* 2016;70:17-25. doi: 10.1016/j.jclinepi.2015.07.002 [published Online First: 2015/07/15]
11. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 2016;3:160018. doi: 10.1038/sdata.2016.18 [published Online First: 2016/03/16]
12. Peters MD, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare* 2015;13(3):141-6. doi: 10.1097/xeb.0000000000000050 [published Online First: 2015/07/03]
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of internal medicine* 2018;169(7):467-73. doi: 10.7326/m18-0850 [published Online First: 2018/09/05]

14. Clinicaltrials.gov. Learn About Clinical Studies [Available from: https://clinicaltrials.gov/ct2/about-studies/learn accessed 2019, november 21th.
15. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of clinical epidemiology* 2016;75:40-6. doi: 10.1016/j.jclinepi.2016.01.021 [published Online First: 2016/03/24]
16. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ (Clinical research ed)* 2012;345:e7570. doi: 10.1136/bmj.e7570 [published Online First: 2012/11/22]
17. Rathi VK, Strait KM, Gross CP, et al. Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials* 2014;15:384. doi: 10.1186/1745-6215-15-384 [published Online First: 2014/10/04]
18. Tannenbaum S, Ross JS, Krumholz HM, et al. Early Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators. *Annals of internal medicine* 2018;169(8):586-88. doi: 10.7326/m18-0723 [published Online First: 2018/07/18]
19. Ji Y, Chen L, Hao X, Wang N & Bao Y. Endorsement of Data Sharing by Authors and High-Impact Medical Journals in China: A Survey of Authors and Assessment of Journal Online Instructions. The 8th International Congress on Peer Review and Biomedical Publication; 2017; Chicago.
20. Griswold M, Laine C, Mulrow C, Schaeffer MB, Stack, C. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. The 6th International Congress on Peer Review and Biomedical Publication; 2009; Vancouver.
21. Rowhani-Farid A, Barnett AG. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ open* 2016;6(10):e011784. doi: 10.1136/bmjopen-2016-011784 [published Online First: 2016/10/16]
22. Murugiah K, Ritchie JD, Desai NR, et al. Availability of Clinical Trial Data From Industry-Sponsored Cardiovascular Trials. *Journal of the American Heart Association* 2016;5(4):e003307. doi: 10.1161/jaha.116.003307 [published Online First: 2016/04/22]
23. Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *Journal of clinical epidemiology* 2019;109:111-16. doi: 10.1016/j.jclinepi.2019.01.012 [published Online First: 2019/02/10]
24. Kaufmann I, Gattrell WT. Assessment of journal compliance with data sharing guidelines from the International Committee of Medical Journal Editors (ICMJE) / in Original Abstracts from the 2019 European Meeting of ISMPP. *Current Medical Research and Opinion* 2019;35(sup2):31-41. doi: 10.1080/03007995.2019.1583496
25. Papageorgiou SN, Antonoglou GN, Martin C, et al. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. *Journal of orthodontics* 2019;46(2):101-09. doi: 10.1177/1465312519842315 [published Online First: 2019/05/09]
26. Kemper JM, Rolnik DL, Mol BWJ, et al. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. *Fertility and sterility* 2020;114(6):1322-29. doi: 10.1016/j.fertnstert.2020.05.020 [published Online First: 2020/08/11]
27. Johnson AL, Torgerson T, Skinner M, et al. An assessment of transparency and reproducibility-related research practices in otolaryngology. *The Laryngoscope* 2020;130(8):1894-901. doi: 10.1002/lary.28322 [published Online First: 2019/10/09]

28. Statham EE, White SA, Sonwane B, et al. Primed to comply: Individual participant data sharing statements on ClinicalTrials.gov. *PloS one* 2020;15(2):e0226143. doi: 10.1371/journal.pone.0226143 [published Online First: 2020/02/19]
29. Mayer CS, Williams N, Gold S, et al. Sharing of Individual Participant Data from Clinical Trials: General Comparison and HIV Use Case. *AMIA Annual Symposium proceedings AMIA Symposium* 2019;2019:647-54. [published Online First: 2020/04/21]
30. Ali J, Califf R, Sugarman J. Anticipated Ethics and Regulatory Challenges in PCORnet: The National Patient-Centered Clinical Research Network. *Accountability in research* 2016;23(2):79-96. doi: 10.1080/08989621.2015.1023951 [published Online First: 2015/07/21]
31. Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15(2):125-34. doi: 10.1111/1467-8519.00220 [published Online First: 2001/11/08]
32. Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *Journal of clinical epidemiology* 2019;106:60-69. doi: 10.1016/j.jclinepi.2018.10.006 [published Online First: 2018/10/22]
33. Godolphin PJ, Bath PM, Montgomery AA. Short email with attachment versus long email without attachment when contacting authors to request unpublished data for a systematic review: a nested randomised trial. *BMJ open* 2019;9(1):e025273. doi: 10.1136/bmjopen-2018-025273 [published Online First: 2019/02/02]
34. Howe N, Giles E, Newbury-Birch D, et al. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *Journal of health services research & policy* 2018;23(2):123-33. doi: 10.1177/1355819617751555 [published Online First: 2018/04/15]
35. Merson L, Phong TV, Nhan le NT, et al. Trust, Respect, and Reciprocity: Informing Culturally Appropriate Data-Sharing Practice in Vietnam. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):251-63. doi: 10.1177/1556264615592387 [published Online First: 2015/08/25]
36. Cheah PY, Tangseefa D, Somsaman A, et al. Perceived Benefits, Harms, and Views About How to Share Data Responsibly: A Qualitative Study of Experiences With and Attitudes Toward Data Sharing Among Research Staff and Community Representatives in Thailand. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):278-89. doi: 10.1177/1556264615592388 [published Online First: 2015/08/25]
37. Broes S, Verbaanderd C, Casteels M, et al. Sharing of Clinical Trial Data and Samples: The Cancer Patient Perspective. *Frontiers in medicine* 2020;7:33. doi: 10.3389/fmed.2020.00033 [published Online First: 2020/03/03]
38. Jones CW, Roberts BW, Platts-Mills TF. Patient Perspectives on Sharing Deidentified Trial Data. *Annals of internal medicine* 2016;165(10):748-49. doi: 10.7326/l16-0089 [published Online First: 2016/11/15]
39. Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *The New England journal of medicine* 2018;378(23):2202-11. doi: 10.1056/NEJMsa1713258 [published Online First: 2018/06/07]
40. Colombo C, Roberto A, Krleza-Jeric K, et al. Sharing individual participant data from clinical studies: a cross-sectional online survey among Italian patient and citizen

groups. *BMJ open* 2019;9(2):e024863. doi: 10.1136/bmjopen-2018-024863 [published Online First: 2019/02/21]
41. Spence O, Onwuchekwa Uba R, Shin S, et al. Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials* 2018;19(1):269. doi: 10.1186/s13063-018-2651-2 [published Online First: 2018/05/05]
42. Krleža-Jerić K, Sim I, Marušić A, Reveiz L & Granados C. Medical Journal Editor Perspectives on Sharing Results Data From Patient-Oriented Research: A WAME Survey. 7th Peer Review Congress; 2009; Vancouver.
43. Chapman SJ, Shelton B, Mahmood H, Bhangu A, Fitzgerald JEF, Harrison EM. Promoting transparency in clinical research: Systematic review of disclosure and data-sharing policies in surgical journals. ASiT; 2014.
44. Chikramane A, Venkatraman L, Khanna V, Sinha S. Data sharing trends and integrity in the era of open science: a qualitative analysis of journals in top-ranked therapeutic areas. 13th Annual Meeting of ISMPP; 2017.
45. Vidal-Infer A, Tarazona B, Alonso-Arroyo A, et al. Public availability of research data in dentistry journals indexed in Journal Citation Reports. *Clinical oral investigations* 2018;22(1):275-80. doi: 10.1007/s00784-017-2108-0 [published Online First: 2017/03/28]
46. Vassar M, Jellison S, Wendelbo H, et al. Data sharing practices in randomized trials of addiction interventions. *Addictive behaviors* 2020;102:106193. doi: 10.1016/j.addbeh.2019.106193 [published Online First: 2019/11/27]
47. Nutu D, Gentili C, Naudet F, et al. Open science practices in clinical psychology journals: An audit study. *Journal of abnormal psychology* 2019;128(6):510-16. doi: 10.1037/abn0000414 [published Online First: 2019/08/02]
48. Gorman DM. Use of publication procedures to improve research integrity by addiction journals. *Addiction (Abingdon, England)* 2019;114(8):1478-86. doi: 10.1111/add.14604 [published Online First: 2019/03/10]
49. Gorman DM, Ferdinand AO. High impact nutrition and dietetics journals' use of publication procedures to increase research transparency. *Research Integrity and Peer Review* 2020;5(1):12. doi: 10.1186/s41073-020-00098-9
50. Almaqrami BS, Hua F, Liu Y, et al. Research waste-related editorial policies of leading dental journals: Situation 2018. *Oral diseases* 2020;26(3):696-706. doi: 10.1111/odi.13257 [published Online First: 2019/12/18]
51. Siebert M, Gaba JF, Caquelin L, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. *BMJ open* 2020;10(5):e038887. doi: 10.1136/bmjopen-2020-038887 [published Online First: 2020/06/01]
52. Aleixandre-Benavent R, Sapena AF, Ferrer SC, et al. Policies regarding public availability of published research data in pediatrics journals. *Scientometrics* 2019;118(2):439-51. doi: 10.1007/s11192-018-2978-1
53. DeVito NJ, French L, Goldacre B. Noncommercial Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *Jama* 2018;319(16):1721-23. doi: 10.1001/jama.2018.2841 [published Online First: 2018/05/02]
54. Whitlock EP, Dunham KM, DiGioia K, et al. Noncommercial US Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability.

*JAMA network open* 2019;2(1):e187498. doi: 10.1001/jamanetworkopen.2018.7498 [published Online First: 2019/01/27]
55. Rollando P, Parc C, Naudet F, et al. [Data sharing policies of clinical trials funders in France]. *Therapie* 2020;75(6):527-36. doi: 10.1016/j.therap.2020.04.001 [published Online First: 2020/05/25]
56. Goldacre B, Lane S, Mahtani KR, et al. Pharmaceutical companies' policies on access to trial data, results, and methods: audit study. *BMJ (Clinical research ed)* 2017;358:j3334. doi: 10.1136/bmj.j3334 [published Online First: 2017/07/28]
57. Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. *BMC medicine* 2018;16(1):165. doi: 10.1186/s12916-018-1154-z [published Online First: 2018/09/29]
58. Gaba JF, Siebert M, Dupuy A, et al. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PloS one* 2020;15(8):e0237464. doi: 10.1371/journal.pone.0237464 [published Online First: 2020/08/21]
59. Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *The New England journal of medicine* 2005;352(21):2202-10. doi: 10.1056/NEJMsa044115 [published Online First: 2005/05/27]
60. Bosserdt M, Hamm B, Dewey M. Clinical trials in radiology and data sharing: results from a survey of the European Society of Radiology (ESR) research committee. *European radiology* 2019;29(9):4794-802. doi: 10.1007/s00330-019-06105-y [published Online First: 2019/02/28]
61. Rowhani-Farid A, Aldcroft A, Barnett AG. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Royal Society open science* 2020;7(3):191818. doi: 10.1098/rsos.191818 [published Online First: 2020/04/10]
62. Veroniki AA, Ashoor HM, Le SPC, et al. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *Journal of clinical epidemiology* 2019;113:176-88. doi: 10.1016/j.jclinepi.2019.05.031 [published Online First: 2019/06/04]
63. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ (Clinical research ed)* 2012;344:d7762. doi: 10.1136/bmj.d7762 [published Online First: 2012/01/05]
64. Nevitt SJ, Marson AG, Davie B, et al. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ (Clinical research ed)* 2017;357:j1390. doi: 10.1136/bmj.j1390 [published Online First: 2017/04/07]
65. Geifman N, Bollyky J, Bhattacharya S, et al. Opening clinical trial data: are the voluntary data-sharing portals enough? *BMC medicine* 2015;13:280. doi: 10.1186/s12916-015-0525-y [published Online First: 2015/11/13]
66. Kochhar S, Knoppers B, Gamble C, et al. Clinical trial data sharing: here's the challenge. *BMJ open* 2019;9(8):e032334. doi: 10.1136/bmjopen-2019-032334 [published Online First: 2019/08/24]
67. Schmidt H, Caldwell B, Mollet A, et al. An Industry Experience with Data Sharing. *The New England journal of medicine* 2018;379(21):2081-82. doi: 10.1056/NEJMc1805610 [published Online First: 2018/11/22]

68. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Scientific data* 2018;5:180268. doi: 10.1038/sdata.2018.268 [published Online First: 2018/11/28]
69. So D, Knoppers BM. Ethics approval in applications for open-access clinical trial data: An analysis of researcher statements to clinicalstudydatarequest.com. *PloS one* 2017;12(9):e0184491. doi: 10.1371/journal.pone.0184491 [published Online First: 2017/09/09]
70. Boutron I, Dechartres A, Baron G, et al. Sharing of Data From Industry-Funded Registered Clinical Trials. *Jama* 2016;315(24):2729-30. doi: 10.1001/jama.2016.6310 [published Online First: 2016/07/02]
71. Miller J, Ross JS, Wilenzick M, et al. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *BMJ (Clinical research ed)* 2019;366:l4217. doi: 10.1136/bmj.l4217 [published Online First: 2019/07/12]
72. Coady SA, Mensah GA, Wagner EL, et al. Use of the National Heart, Lung, and Blood Institute Data Repository. *The New England journal of medicine* 2017;376(19):1849-58. doi: 10.1056/NEJMsa1603542 [published Online First: 2017/04/13]
73. Shmueli-Blumberg D, Hu L, Allen C, et al. The national drug abuse treatment clinical trials network data share project: website design, usage, challenges, and future directions. *Clinical trials (London, England)* 2013;10(6):977-86. doi: 10.1177/1740774513503522 [published Online First: 2013/10/03]
74. Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. *Clinical trials (London, England)* 2018;15(4):413-23. doi: 10.1177/1740774518769655 [published Online First: 2018/04/21]
75. Bertagnolli MM, Sartor O, Chabner BA, et al. Advantages of a Truly Open-Access Data-Sharing Model. *The New England journal of medicine* 2017;376(12):1178-81. doi: 10.1056/NEJMsb1702054 [published Online First: 2017/03/23]
76. Asare AL, Carey VJ, Rotrosen D, et al. Clinical trial data access: Opening doors with TrialShare. *The Journal of allergy and clinical immunology* 2016;138(3):724-26. doi: 10.1016/j.jaci.2016.05.034 [published Online First: 2016/07/31]
77. MRC. MRC Clinical Trials Transparency Review 2017 [Available from: https://mrc.ukri.org/documents/pdf/clinical-trials-review-2017/ accessed 22/11/2019.
78. Zhu CS, Pinsky PF, Moler JE, et al. Data sharing in clinical trials: An experience with two large cancer screening trials. *PLoS medicine* 2017;14(5):e1002304. doi: 10.1371/journal.pmed.1002304 [published Online First: 2017/05/26]
79. Chalmers I, Altman DG, McHaffie H, et al. Data sharing among data monitoring committees and responsibilities to patients and science. *Trials* 2013;14:102. doi: 10.1186/1745-6215-14-102 [published Online First: 2013/06/21]
80. Tudur Smith C, Nevitt S, Appelbe D, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials* 2017;18(1):319. doi: 10.1186/s13063-017-2067-4 [published Online First: 2017/07/18]
81. Strom BL, Buyse M, Hughes J, et al. Data sharing, year 1--access to data from industry-sponsored clinical trials. *The New England journal of medicine* 2014;371(22):2052-4. doi: 10.1056/NEJMp1411794 [published Online First: 2014/10/16]

82. Strom BL, Buyse ME, Hughes J, et al. Data Sharing - Is the Juice Worth the Squeeze? *The New England journal of medicine* 2016;375(17):1608-09. doi: 10.1056/NEJMp1610336 [published Online First: 2016/10/27]
83. Ross JS, Ritchie JD, Bamford S, Berlin JA, Karla Childers K, Desai N, Gamble GM, Gross CP, Richard S, Lehman RS, Lins P, Morris S, Waldstreicher J, Krumholz H. Sharing Data Through the Yale University Open Data Access (YODA) Project: Early Experience. 8th Peer Review Congress; 2017; Chicago.
84. Lovato L. Exploring the potential of data sharing: The SPRINT Challenge in Abstracts from the SCT 39th Annual Meeting (2018). *Clinical trials (London, England)* 2018;15(2_suppl):S35-s192. doi: 10.1177/1740774518790846 [published Online First: 2018/08/22]
85. Kawahara T, Fukuda M, Oba K, et al. Meta-analysis of randomized clinical trials in the era of individual patient data sharing. *International journal of clinical oncology* 2018;23(3):403-09. doi: 10.1007/s10147-018-1237-z [published Online First: 2018/01/14]
86. Ebrahim S, Sohani ZN, Montoya L, et al. Reanalyses of randomized clinical trial data. *Jama* 2014;312(10):1024-32. doi: 10.1001/jama.2014.9646 [published Online First: 2014/09/10]
87. Naudet F, Sakarovitch C, Janiaud P, 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 The BMJ and PLOS Medicine. *BMJ (Clinical research ed)* 2018;360:k400. doi: 10.1136/bmj.k400 [published Online First: 2018/02/15]
88. Ross JS, Ritchie JD, Finn E, et al. Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ open* 2016;6(9):e012769. doi: 10.1136/bmjopen-2016-012769 [published Online First: 2016/09/28]
89. Vaduganathan M, Nagarur A, Qamar A, et al. Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation* 2018;137(9):938-47. doi: 10.1161/circulationaha.117.031883 [published Online First: 2017/11/15]
90. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PloS one* 2007;2(3):e308. doi: 10.1371/journal.pone.0000308 [published Online First: 2007/03/22]
91. ECRIN. ECRIN metadata repository [Available from: https://ecrin.org/tools/clinical-research-metadata-repository accessed 01.07.2021 2021.
92. Doshi P, Shamseer L, Jones M, et al. Restoring biomedical literature with RIAT. *Bmj* 2018;361:k1742. doi: 10.1136/bmj.k1742 [published Online First: 2018/04/28]
93. Takata J, Sondo P, Humphreys GS, et al. The WorldWide Antimalarial Resistance Network Clinical Trials Publication Library: A Live, Open-Access Database of Plasmodium Treatment Efficacy Trials. *The American journal of tropical medicine and hygiene* 2020;103(1):359-68. doi: 10.4269/ajtmh.19-0706 [published Online First: 2020/05/21]
94. Crossmark [Available from: https://www.crossref.org/services/crossmark/ accessed 01.11.2021 2021.
95. Devereaux PJ, Guyatt G, Gerstein H, et al. Toward Fairness in Data Sharing. *The New England journal of medicine* 2016;375(5):405-7. doi: 10.1056/NEJMp1605654 [published Online First: 2016/08/16]

96. Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic reviews* 2012;1:2. doi: 10.1186/2046-4053-1-2 [published Online First: 2012/05/17]

**ABBREVIATIONS**

ANR: Agence Nationale de la Recherche
ASAPbio: Accelerating Science and Publication in biology
BioLINCC: Biological Specimen and Data Repository Information Coordination Center
CDAS: Cancer Data Access System
CIHR: Canadian Institutes of Health Research
CSDR: Clinical Study Data Request
CT.gov: ClinicalTrials.gov
CTN: Clinical Trials Network
DFG: Deutsche Forschungsgemeinschaft
DGOS : Direction Générale de l'Offre de Soins
Drum: Data Repository for University of Minnesota
EBCTG: Early Breast Cancer Trialists' Collaborative Group
EC Europe: European Commission
EFPIA: European Federation of Pharmaceutical Industries and Associations
F1000Research: Faculty of 1000 Research
FAIR: Findable, Accessible, Interoperable and Reusable
ICMJE: International Committee of Medical Journal Editors
ICPSR: Inter-university Consortium for Political and Social Research
IOM: Institute Of Medicine
IPD: Individual Participant Data
ITN Trialshare: Immune Tolerance Network TrialShare
MMMP: Melanoma Molecular Map Project
MRC UK: Medical Research Council
NHMRC: National Health and Medical Research Council
NIDA: National Institute on Drug Abuse
NIDDK: National institute of Diabetes and Digestive and Kidney Diseases
NIH: National Institute of Health
NIH BioLINCC: National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center
NIHR: National Institute of Health Research
NIMH NDCT: National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness
NSFC: National Natural Science Foundation of China
PCORNeT: The National Patient-centered Clinical research Network
PHRC: Le programme hospitalier de recherche clinique
PHRMA: Pharmaceutical Research and Manufacturers of America
PI: Principal Investigator
PLOS: Public Library Of Science
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: extension for Scoping Reviews
ProAct: Pooled Resource Open Access Clinical trials database
RCT: Randomised clinical trial
RDA: Research Data Alliance
SOAR: the Supporting Open Access for Researchers initiative
SND: Swedish National Data Service
TBI-IMPACT: Traumatic Brain Injury– International Mission for Prognosis and Analysis of Clinical trials in Traumatic brain injury
The BMJ: The British Medical Journal
UK: United Kingdom
UKCRC: UK Clinical Research Collaboration
UMIN: University Medical Hospital Information Network
US: United States of America
US DoD: United States Department of Defense
ViVli: adapted from the Greek "ViVliothiki" (library) and the Latin root "viv" (life)
WWARN: World Wide Antimalarial Resistance Network
YODA: the Yale University Open Data Access Project

**FIGURES**

**Figure 1**: PRISMA flow diagram (PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses).
** For National Institute of Health (NIH US), the answer we received was not informative

**Figure 2**: Proportion of the 93 references exploring each outcome domain

**Figure 3**: Outcomes used to assess current data sharing practices of individual patient data for clinical trials organized per outcome domains and number of studies exploring these outcomes
. Experimental = experimental (i.e. randomised) studies comparing prospectively at least two interventions
. Survey = surveys, for instance of authors, policies
. Metrics = metrics of data use
. Qualitative = qualitative research
. Other = any other type of research such as case studies for instance

**Figure 4**: Intent to share

a: This rate is 73 % if the purpose is a re-analysis

b: These are 54 participants of 60 who had a opinion about data sharing (other had no knowledge or no opinion)

c: An additional 25 % are undecided

d: This rate is 19 % for requiring a data sharing plan

e: 35 % have a data sharing policy (encouraging data sharing)

f: Only 2 with a mandatory. policy

g: This rate is 71 for a sample of all companies (not only the top 25)

**Figure 5**: Actual data sharing

**Figure 6**: Temporal trends, number and type of published output from CSDR and YODA
Red colour = Studies from CSDR
Blue colour = Studies from YODA

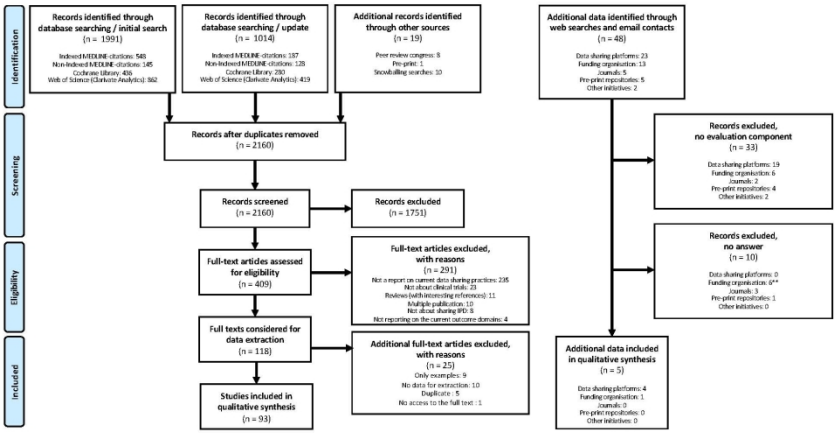
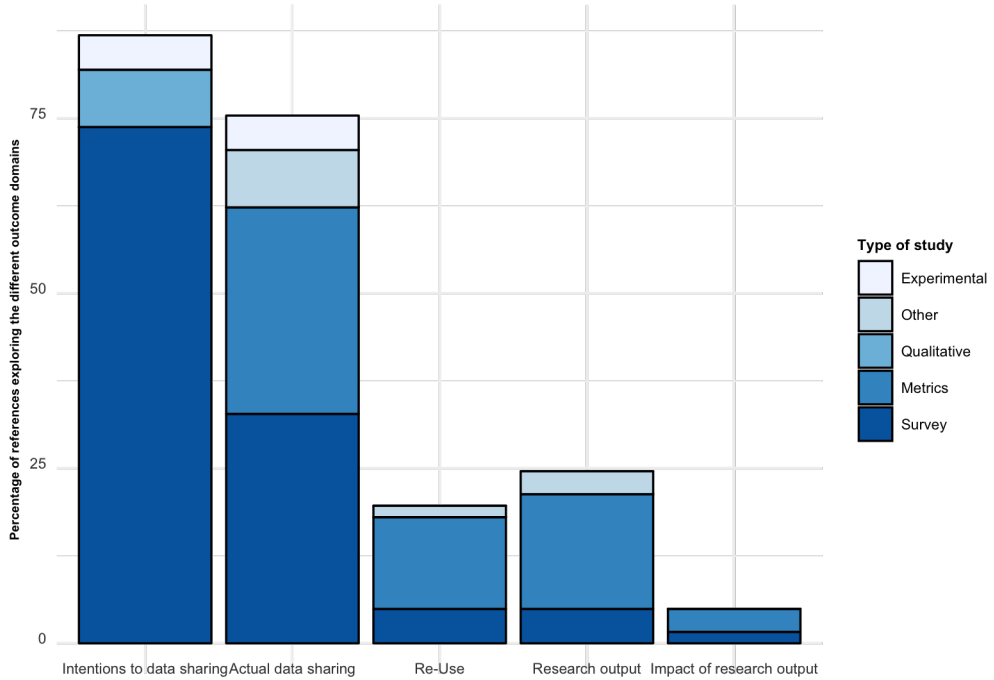
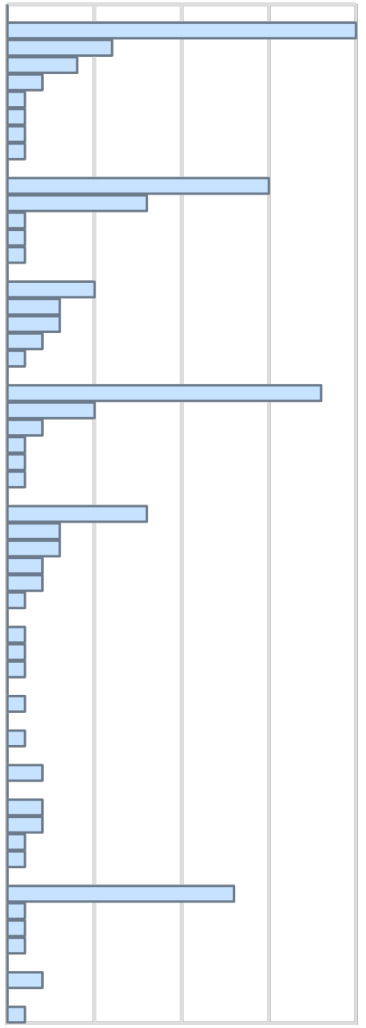



Figure 1: PRISMA flow diagram (PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses).
** For National Institute of Health (NIH US), the answer we received was not informative

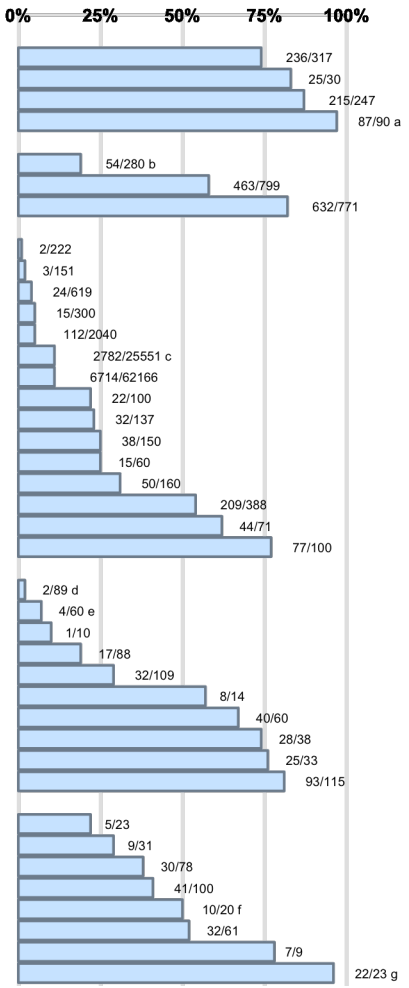
297x209mm (200 x 200 DPI)


Percentage of references exploring the different outcome domains
75
50
25
0
Intentions to data sharing
Actual data sharing
Re-Use
Research output
Impact of research output
Type of study
Experimental
Other
Qualitative
Metrics
Survey

| Outcome domains | Outcomes | Experimental | Survey | Metrics | Qualitative | Other |
|---|---|---|---|---|---|---|
| Intentions to data sharing | **For trialists** | . | . | . | . | . |
| | Intentions to share data | 2 | 18 | . | . | . |
| | Existence of a data sharing plan/data sharing statement | . | 6 | . | . | . |
| | Type of data sharing plan | . | 4 | . | . | . |
| | Support for data sharing | . | 2 | . | . | . |
| | Willingness to store IPD in a central repository, conditions for storage | . | 1 | . | . | . |
| | Most pressing ethical issues | . | 1 | . | . | . |
| | Response to a request for data sharing and time to response | 1 | . | . | . | . |
| | Preparedness for DS | 1 | . | . | . | . |
| | **For publishers/funders** | . | . | . | . | . |
| | Existence of a data sharing policy/intent to share | . | 15 | . | . | . |
| | Type of data sharing policy | . | 8 | . | . | . |
| | Journals requiring participant level data | . | 1 | . | . | . |
| | Prohibiting data sharing in agreements with industry | . | 1 | . | . | . |
| | Existence of a data sharing policy for clinical trial units | . | 1 | . | . | . |
| | **For trial participants** | . | . | . | . | . |
| | Willingness to share data | . | 4 | . | 1 | . |
| | Barriers to share data | . | 2 | . | 1 | . |
| | Patient Opinions on the Release of Deidentified Individual-Patient Data | . | 3 | . | . | . |
| | Views, experience and attitudes towards DS | . | . | . | 2 | . |
| | Information in consent forms | . | 1 | . | . | . |
| Actual data sharing | **For re-users (e.g. IPD meta-analysts)** | . | . | . | . | . |
| | Data availability (yes/no) | 3 | 13 | . | . | 2 |
| | Data release after a request | . | 4 | 1 | . | . |
| | Data availability (time) | 1 | . | 1 | . | . |
| | Data availability (completeness) | 1 | . | . | . | . |
| | Rate of IPD meta-analyses / all requests | . | . | 1 | . | . |
| | Reasons for request, research plan and experience of DS | . | 1 | . | . | . |
| | **For repositories/platforms** | . | . | . | . | . |
| | Data release after a request | . | 2 | 5 | . | 1 |
| | Approval of data release | . | 2 | 1 | . | . |
| | Data availability | . | 3 | . | . | . |
| | Number of downloads | . | . | 2 | . | . |
| | Number of DS agreement, speed of data availability | . | 1 | 1 | . | . |
| | Database access | . | . | 1 | . | . |
| | **For other stakeholders** | . | . | . | . | . |
| | Number of downloads and page hits for any public access | . | . | 1 | . | . |
| | Re-use of data by independent data monitoring committees | . | . | . | . | 1 |
| | Cost of preparing the data for data generators | . | . | . | . | 1 |
| Re-use | **Further additional analyses** | . | . | . | . | . |
| | Approved proposals by subject areas | . | . | 1 | . | . |
| | **IPD meta-analyses** | . | . | . | . | . |
| | Proposals for IPD-meta-analyses and one concrete example | . | . | 1 | . | . |
| | **Re-analyses** | . | . | . | . | . |
| | Reproducibility on primary outcomes | . | 2 | . | . | . |
| | **Any type of re-use** | . | . | . | . | . |
| | Type of re-use | . | . | 2 | . | . |
| | Progression of the analysis | . | . | 1 | . | 1 |
| | Listing of type of studies performed | . | . | 1 | . | . |
| | Project completion | . | 1 | . | . | . |
| Output from data sharing | **Published re-use** | . | . | . | . | . |
| | Publication of re-use (papers) | . | 2 | 9 | . | 2 |
| | Manuscripts in peer review | . | . | 1 | . | . |
| | Request discontinued | . | . | 1 | . | . |
| | Communication of re-use (oral presentation, posters) | . | . | . | . | 1 |
| Impact of research output | **Quantitative metrics for published articles** | . | . | . | . | . |
| | Citation metrics | . | 1 | 1 | . | . |
| | **Identification of a new finding** | . | . | . | . | . |
| | Crowd voting | . | 1 | . | . | . |



| Reference | Type | Population | Time point |
|---|---|---|---|
| **A: Intent to share: surveys of trialists** | | | |
| **Rathi, 2012** | Overall | Trials published in 6 high impact journals | 2010-2011 |
| **Tudur-Smith, 2014** | Overall (central storage of their IPD) | Reviewers of Cochrane IPD meta-analysis group | 2011 |
| **Yuanyuan, 2017** | Overall (endorsment of DS) | Trials publihed in Chinese Medical Journal | 2016 |
| **Tannenbaum, 2018** | For an IPD meta-analysis | Trials published in 3 high impact journals with data sharing policies | 2012-2016 |
| **B: Intent to share: surveys of trial participants** | | | |
| **Colombo, 2017** | Having knowlegde about and being in favor of data sharing | Italian patients and citizen groups | 2017 |
| **Jones, 2016** | Favor or strongly favor data sharing | Patients in a Usemergency department | 2015 |
| **Mello, 2018** | Perception that the benefits of data sharing outweighed the negative aspects | Patients from 3 US medical centers | Unclear |
| **C: Intent to share: data sharing statements** | | | |
| **Kemper, 2020** | Overall | Reproductive endocrinology and infertility articles (study mix) | 2013, 2018 |
| **Johnson, 2020** | Overall | 300 otolaryngology research studies | 2014-2018 |
| **Gabelica, 2019** | Overall | RCTs in 7 high-ranked anesthesiology journals | 2014-2016 |
| **Papageorgios, 2019** | Open data | Trials in orthodontics and periodontics | 2017-2018 |
| **Statham, 2020** | Overall | CT.gov | 2018 |
| **Bergeris, 2018** | Overall | CT.gov | Up to august 2017 |
| **Mayer, 2019** | Studies with a data sharing plan | CT.gov | 2015-2018 |
| **Siebert, 2020b** | Overall | ICMJE affiliates (after policy) | 2019 |
| **Kaufman, 2019** | Overall | RCTs in 11 selected journals (before policy) | 2018 |
| **Kaufman, 2019b** | Overall | RCTs in 11 selected journals (after policy) | 2018 |
| **Murugiah, 2016** | Data made available | Clinical trials (> 5000 patients), from clinicaltrials.gov | Up to 2015 |
| **Rowhani-Farid, 2016** | Overall | The BMJ (all research papers) | 2009-2015 |
| **Griswold M, 2013** | Overall | Ann. Int. Med. (all research papers) | 2008-2012 |
| **Laine, 2009** | Overall | Ann. Int. Med. (all research papers) | 2008 |
| **Siebert, 2020** | Overall | ICMJE members (after policy) | 2019 |
| **D: Intent to share: journal data sharing policies** | | | |
| **Krleza-Jeric, 2009** | Require IPD | Members of World Association of Medical Editors | 2009 |
| **Chickramane, 2017** | Require IPD | 15 oncology, 15 central nervous system, 15 cardiology/endocrinology and 15 respiratory journals | Unknown |
| **Chapman, 2014** | Data sharing policy | High impact surgical journals | 2009-2012 |
| **Vidal-Infer, 2018** | Accept IPD as a complementary material | Dental journals | 2014 |
| **Almaqrami, 2020** | Data sharing policy | Dental journals | 2018 |
| **Vasar, 2020** | Data sharing policy | 15 high-impact addiction journals | 2013-2018 |
| **Nutu, 2019** | Data sharing policy | Clinical psychology journals | 2017 |
| **Gorman, 2019** | Data sharing policy | High impact addiction journals | 2018 |
| **Gorman, 2020** | Data sharing policy | Nutrition and dietetics journals | 2018 |
| **Alexandre-Benavent, 2019** | Data sharing policy | Paediatric journals | 2012-2016 |
| **E: Intent to share: funders and clinical trial units** | | | |
| **Hopkins, 2016** | Data sharing policy | UK CTUs | 2014 |
| **Rollando, 2020** | Data sharing policy | Clinical trial funders in France | 2019 |
| **Gaba, 2020** | Data sharing policy | Non-commercial | 2016-2018 |
| **Gaba, 2020b** | Data sharing policy | Commercial funders | 2016-2018 |
| **de Vito, 2018** | Data sharing requirement | Top non commercial funders | 2017 |
| **Hopkins, 2018** | Data sharing policy | Clinical trials sponsored by the pharmaceutical industry published in the top 10 medical journals | 2015 |
| **Whitlock, 2019** | Data sharing policy | Non-commercial funders in the US | 2018 |
| **Goldacre, 2017** | Data sharing policy | Top 25 pharmaceutical companies by revenue | 2016 |



**A: Actual data sharing by re-users: surveys related to published studies**

| Reference | Data source | Sample selection | Time period |
|---|---|---|---|
| Vassar, 2020 | 15 high ranked addiction journals | Consecutive | 2013-2018 |
| Gabelica, 2019 | 7 high ranked addiction journals | Consecutive | 2014-2016 |
| Savage, 2009 | PLOS Medicine, PLOS clinical trials | Unclear | 2009 |
| Yuanyuan, 2017 | Chinese Medical Journal | Consecutive | 2016 |
| Hopkins, 2018 | Top 10 general and internal medical journals | Consecutive, industry-sponsored trials | 2015 |
| Rowhani-Farid, 2016 | BMJ | Random, subsample RCTs | 2009-2015 |
| Rathi, 2012 | 6 general medical journals | Consecutive | 2010-2011 |
| Naudet, 2018 | BMJ, PLOS Medicine | Consecutive, RCTs | 2013-2016 |
| Tannenbaum, 2018 | BMJ, PLOS Medicine, Ann. Inn. Med. | Selected, partly restricted to authors willing to share data | 2012-2016 |

**B: Actual data sharing by re-users: data related to IPD meta-analyses**

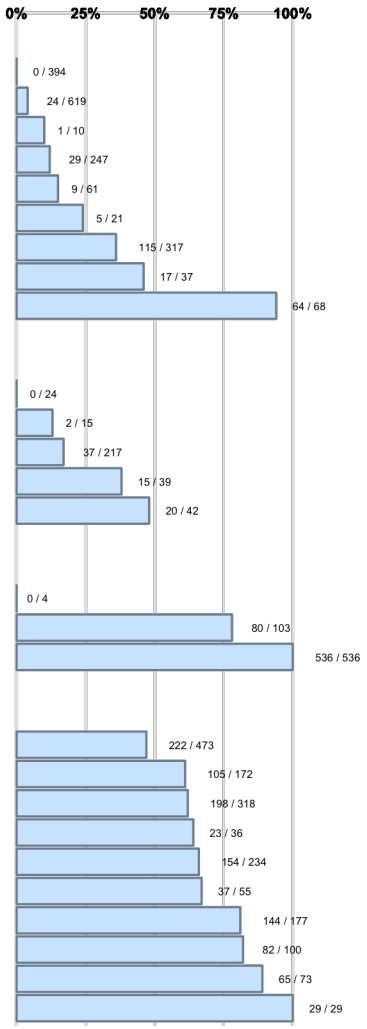
| Reference | Number of studies with IPD sharing | Time point | Comment |
|---|---|---|---|
| Mayo-Wilson, 2015 | 0/24 (0%) | Contacted in 2014 | Commercial sponsor, trials with one medicinal product |
| Kawahara, 2018 | 2/15 (13%) through CSDR | Unknown | 13 requested from authors directly, project in progress |
| Villein, 2015 | 37/217 (17%) | RCTs with results published since 2000 | |
| Nevitt,2017 | 15/35 (38%), 4/15 through CSDR | End 2015 | |
| Hee, 2016 | 20/42 (48%) | RCTs until 2011 | |

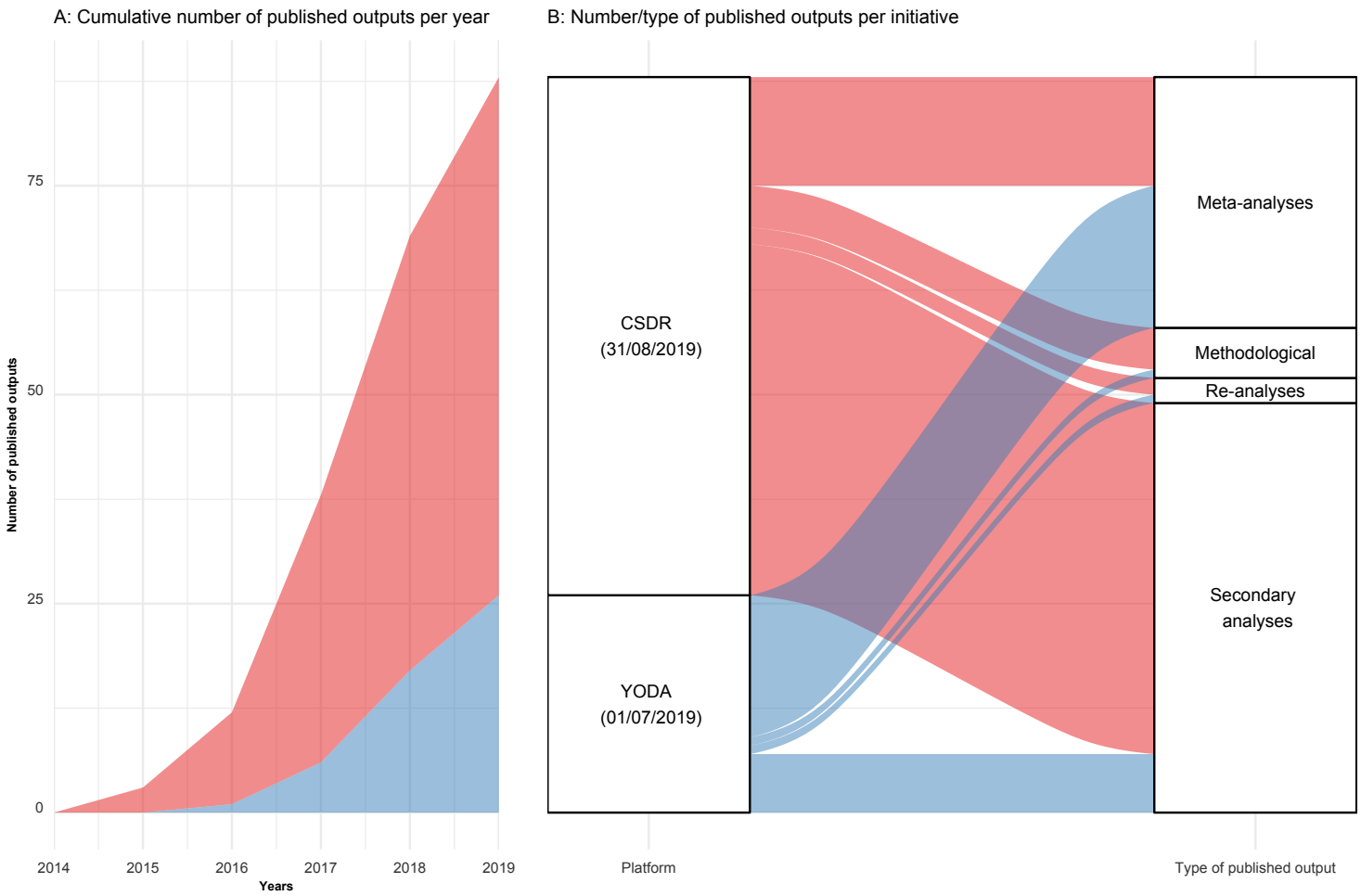
**C: Actual data sharing by re-users: repositories/platforms from the viewpoint of the user**

| Reference | Repository_platform | Time point of submission | Comment |
|---|---|---|---|
| Geifman, 2015 | CSDR | 12/2014-1/2015 | |
| Sydes, 2015 | MRC CTU | 2012-2014 | 103 requests to 54 trials |
| Ross, 2016 | BioLINCC | 2007-2014 | Survey of investigators who received access to BioLINCC |

**D: Actual data sharing by re-users: survey of repositories/platforms**

| Reference | Platform | Time point of assessment | Comment |
|---|---|---|---|
| Kochar, 2019 | CSDR | 2014-2019 | |
| So, 2017 | CSDR | February 2017 | |
| Vaduganathan, 2018 | CSDR | May 2017 | |
| Strom, 2014 | CSDR | May 2014 | |
| Navar, 2016 | CSDR, YODA, SOAR | December 2015 | |
| Schmidt, 2018 | CSDR | 2014-2017 | Boehringer-Ingelheim's studies |
| Strom, 2016 | CSDR | November 2015 | 177 for 237 trials |
| Ross, 2018 | YODA | August 2018 | |
| Ross, 2017 | YODA | June 2017 | 73 for 159 trials |
| Krumholz, 2016 | YODA | 2015 | |




A: Cumulative number of published outputs per year
Number of published outputs
0
25
50
75
2014
2015
2016
2017
2018
2019
Years
B: Number/type of published outputs per initiative
CSDR
(31/08/2019)
YODA
(01/07/2019)
Platform
Meta-analyses
Methodological
Re-analyses
Secondary
analyses
Type of published output

# Supplementary material 1: Information sources

For commercial sponsors, we considered:
1/ Clinical Study Data Request (CSDR),
2/ the Yale University Open Data Access Project (YODA),
3/ the Supporting Open Access for Researchers initiative (SOAR),
4/ ViVli.

For non-commercial sponsor, we considered:
1/ the National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness (NIMH NDCT),
2/ The National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center (NIH BioLINCC),
3/ B2Share,
4/ Dryad,
5/ the Data Repository for University of Minnesota (Drum),
6/ EASY,
7/ Edinburgh DataShare,
8/ FigShare,
9/ the Inter-university Consortium for Political and Social Research (ICPSR),
10/ the Swedish National data Service (SND),
11/ the University Medical Hospital Information Network (UMIN),
12/ Zenodo,
13/ the Early Breast Cancer Trialists' Collaborative Group (EBCTG),
14/ FreeBird,
15/ Traumatic Brain Injury – International Mission for Prognosis and Analysis of Clinical trials in TBI (TBI-IMPACT),
16/ Melanoma Molecular Map Project (MMMP),
17/ National institute of Diabetes and Digestive and Kidney Diseases (NIDDK),
18/ Immune Tolerance Network TrialShare (ITN Trialshare),
19/ Child Abuse,
20/ Pooled Resource Open Access Clinical trials database (ProAct).

For the different funders:
1/ National Institute of Health (NIH US),
2/ European Commission (EC Europe),
3/ Medical Research Council (MRC UK),
3/ Le programme hospitalier de recherche clinique (DGOS France),
4/ L'Agence nationale de la recherche (ANR France),
5/ Department of Defense (US DoD),
6/ Wellcome Trust UK,

7/ Canadian Institutes of Health Research (CIHR Canada),
8/ National Health and Medical Research Council (NHMRC Australia),
9/ Deutsche Forschungsgemeinschaft (DFG Germany),
10/ National Natural Science Foundation of China (NSFC China),
11/ National Institute of Health Research (NIHR UK),
12/ Gates Foundation US.

# Supplementary material 2: Literature searches

The initial algorithm for the literature search, detailed in the registered protocol (osf.io/pb8cj), was updated on October 29th 2018 to include broader search terms.

| **Name of Database** | **Host, search interface** | **Initial search** Search date 2018-10-29 | | **Update search** Search date 2020-09-11 Publication year 2018-2020 | |
|---|---|---|---|---|---|
| | | **Update status of the database** | **Results** | **Update status of the database** | **Results** Publication year From 2018-2020 |
| Medline | Wolters Kluwer / Ovid | 1946 to October Week 3 2018 | 548 | 1946 to September Week 1 2020 | 187 |
| Medline daily update | | October 25, 2018 | | September 09, 2020 | |
| MEDLINE In-Process & Other Non-Indexed Citations | | | 145 | 1946 to September 09, 2020 | 128 |
| MEDLINE Epub Ahead of Print | | | | September 09, 2020 | |
| Cochrane Library: Cochrane Reviews | Wiley | Issue 10 of 12, October 2018 | 19 | Issue 9 of 12, September 2020 | 12 |
| Cochrane Protocols | | | 1 | | 0 |
| Cochrane Central Register of Controlled Trials | | Issue 9 of 12, September 2018 | 416 | Issue 9 of 12, September 2020 | 268 |
| Science Citation Index | Clarivate Analytics / Web of Science | 1945 –present (2018-10-26) | 862 | 2018 –present (2020-09-10) | 419 |
| Social Science Citation Index | | 1956-present (2018-10-26) | | | |

| | | | | |
|---|---|---|---|---|
| **Total with duplicates** | | **1991** | | **1014** |
| **Total without duplicates** | | **1544** | | **763** |
| **New citations 2018-2020 without overlap from initial search** | | | | **597** |
| **Total without overlap initial search and update search (see PRISMA flow diagram)** | | **2141 (1544 + 597)** | | |

**MEDLINE Databases: Host: Wolters Kluwer, search interface: Ovid**

**1. Indexed MEDLINE-citations:**
Search Strategy:

| # | **Searches** | **Results Initial search** Search date: 2018-10-29: MEDLINE 1946 to October Week 3 2018, MEDLINE Daily Update October 25, 2018 | **Results Update search** Search date: 2020-09-11: MEDLINE 1946 to September Week 1 2020, MEDLINE (Daily Update September 09, 2020. | **Annotations** |
|---|---|---|---|---|
| 1 | exp Access to Information/ | 6845 | 7597 | Concept data sharing: MeSH terms |
| 2 | Information Dissemination/ | 14697 | 16894 | |
| 3 | exp *"Information Storage and Retrieval"/ | 52187 | 58658 | |
| 4 | data collection/ | 87165 | 89553 | |
| 5 | datasets as topic/ | 2259 | 4417 | |
| 6 | or/1-5 | 157717 | 170739 | |
| 7 | exp clinical trial/ | 809623 | 868410 | Concept clinical trials: |

| | | | | |
|---|---|---|---|---|
| 8 | exp clinical trial as topic/ | 318580 | 345552 | MeSH terms or textwords |
| 9 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 879338 | 997736 | |
| 10 | Meta-Analysis as Topic/ | 16485 | 18279 | Concept Meta-analysis: MeSH |
| 11 | meta-analysis/ | 93492 | 119228 | |
| 12 | or/7-11 | 1489253 | 1650537 | Concept Clinical Trials OR Meta-analysis |
| 13 | 6 and 12 | 10206 | 11264 | Combination of concepts: data sharing (MeSH only) AND (clinical trials OR meta-analysis) |
| 14 | (data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 7822 | 9921 | Concept data sharing: Textwords |
| 15 | 13 and 14 | 325 | 422 | data sharing (MeSH terms) AND data sharing (textwords) AND (clinical trials OR meta-analysis):<br>1. interim result |
| 16 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 756 | 993 | Concept sharing IPD (textwords) |
| 17 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 23 | 38 | |
| 18 | or/16-17 | 772 | 1017 | |
| 19 | 12 and 18 | 129 | 179 | (Clinical trials OR meta-analysis) AND textwords for sharing IPD:<br>2. interim result |
| 20 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3196 | 4262 | Concept data sharing textwords |

| | | | | |
|---|---|---|---|---|
| 21 | 12 and 20 | 393 | 539 | (Clinical trials OR meta-analysis) AND textwords data sharing:<br>3. interim result |
| 22 | 15 or 19 or 21 | 557 | 738 | OR-combination of interim results |
| 23 | exp animals/ not humans/ | 4508403 | 4732433 | Exclusion of animals only |
| 24 | 22 not 23 | 552 | 732 | |
| 25 | limit 24 to (english or french or german or spanish) | 548 | 725 | Restriction to English, German, French, Spanish:<br>Final result for indexed Medline citations |
| | | | 187 | Update search:<br>limit 25 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Exp **term*/ = MeSH as major topic incl. narrower terms as major topic
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters
adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**2. Non-Indexed MEDLINE-citations:**

| # | **Searches** | **Results Initial search**<br>Search date: 2018-10-29:<br>MEDLINE In-Process & Other Non-Indexed | **Results Update search**<br>Search date: 2020-09-11:<br>MEDLINE Epub Ahead of Print September 09, 2020, MEDLINE | **Annotations** |
|---|---|---|---|---|

| | | Citations October 25, 2018, MEDLINE Epub Ahead of Print October 25, 2018 | In-Process & Other Non-Indexed Citations 1946 to September 09, 2020 | |
|---|---|---|---|---|
| 1 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 215 | 323 | Concept data sharing |
| 2 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3 | 7 | |
| 3 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 1122 | 1564 | |
| 4 | or/1-3 | 1230 | 1727 | |
| 5 | exp clinical trial/ | 401 | 521 | Concept clinical trials |
| 6 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 138560 | 174420 | |
| 7 | meta-analysis as topic/ | 1 | 0 | Concept meta-analysis |
| 8 | meta-analysis/ | 34 | 99 | |
| 9 | (meta-analy* or metaanaly*).ti,ab,kf. | 27362 | 37834 | |
| 10 | or/5-9 | 156727 | 199473 | Concept clinical trials OR meta-analysis |
| 11 | 4 and 10 | 146 | 191 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| 12 | limit 11 to (english or french or german or spanish) | 145 | 191 | Restriction to English, French, German, Spanish:<br>Final result for non-indexed Medline citations |
| | | | 128 | Update search:<br>limit 12 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters

adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**Cochrane Library** (Wiley):
- Cochrane Database of Systematic Reviews
- Cochrane Protocols
- Cochrane Central Register of Controlled Trials

| ID | Search | Annotations |
|---|---|---|
| #1 | ((data near share*) or (data near sharing*)):ti,ab,kw | Concept data sharing: Textword search in title, abstract, keywords |
| #2 | (data next share*) or (data next sharing*) | Concept data sharing: Textword search in fulltext |
| #3 | #1 or #2 | Concept data sharing. 1. Interim result |
| #4 | ((patient* or participant*) near individual*):ti,ab,kw | Concept Individual patient data sharing. 2. Interim result |
| #5 | data:ti,ab,kw | |
| #6 | (share* or sharing*):ti,ab,kw | |
| #7 | #4 and #5 and #6 | |
| #8 | (IPD near (share* or sharing*)):ti,ab,kw | Concept IPD sharing: 3. Interim result |
| #9 | #3 or #7 or #8 in Cochrane Reviews, Cochrane Protocols, Trials | OR-combination of interim results. Limit to Cochrane Reviews, Protocols, Trials: final result |

| Results | Initial search<br>Search date 2018-10-29 | Update search<br>Publication Year 2018-2020<br>Search date 2020-09-11 |
|---|---|---|
| Cochrane Reviews | 19<br>Issue 10 of 12, October 2018 | 12<br>Issue 9 of 12, September 2020 |
| Cochrane Protocols | 1<br>Issue 10 of 12, October 2018 | 0<br>Issue 9 of 12, September 2020 |
| Trials | 416<br>Issue 9 of 12, September 2018 | 268<br>Issue 9 of 12, September 2020 |

ti,ab,kw = title,abstract keywords
near = terms in any order (default: within 6 words)
next = phrase searching: terms next to each other in the given order
* = truncation

Via Web of Science (Clarivate Analytics):
Databases:
- **Science Citation Index Expanded** (SCI-EXPANDED): 1945-present
- **Social Sciences Citation Index** (SSCI): 1956-present

| Set | Query | **Results Initial search** Search date: 2018-10-29 **Time span: all years** Data last updated: 2018-10-26 | **Results update search** Search date: 2020-09-11 **Timespan=2018-2020** Data last updated: 2020-09-10 | **Annotations** |
|---|---|---|---|---|
| # 1 | ts=(("data" near/3 share*) or ("data" near/3 sharing*)) | 12,398 | 5,227 | Concept data sharing |
| # 2 | ts=((patient* or participant*) near/3 individual*) | 73,929 | 17,281 | Concept Individual patient data sharing |
| # 3 | ts="data" | 5,101,598 | 1,087,419 | |
| # 4 | ts=(share* or sharing*) | 456,989 | 114,300 | |
| # 5 | #4 AND #3 AND #2 | 714 | 336 | |
| # 6 | ts=("IPD" near/6 (share* or sharing*)) | 23 | 24 | Concept IPD sharing |
| # 7 | #6 OR #5 OR #1 | 12,970 | 5,478 | OR-combination of concepts |
| # 8 | ts=(randomi?ed or "randomly" or randomi?ation) | 1,044,391 | 201,056 | Concept clinical trials |
| # 9 | ts=((random* or "clinical") near/3 trial*) | 773,198 | 154,885 | |
| # 10 | ts=("meta analy*" or metaanaly*) | 313,038 | 105,276 | Concept meta-analysis |
| # 11 | #10 or #9 OR #8 | 1,478,458 | 323,705 | Concept clinical trials OR meta-analysis |
| # 12 | #11 AND #7 | 1,022 | 453 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| # 13 | #11 AND #7 **Refined by: DOCUMENT TYPES:** ( ARTICLE OR REVIEW ) | 862 | 419 | Restriction to Article or Review: final result |

ts = topic: Title, Abstract, Author Keywords, Keywords Plus®
near/n = terms in any order within n words
* = truncation
? = wildcard for exact 1 character

# Supplementary material 3: Study characteristics

| Author | Year | Country | Type of research | Detail if type of research=other | Type of shared material |
|---|---|---|---|---|---|
| **Tudur-Smith C** | 2014 | UK | Survey | | IPD |
| **Murugiah K** | 2016 | US | Survey | | IPD |
| **Krleža-Jerić K** | 2009 | Canada | Survey | | IPD |
| **Jones CW** | 2016 | US | Survey | | IPD |
| **Mayo-Wilson E** | 2015 | US | Other | Case study | IPD |
| **Reidpath DD** | 2001 | Australia | Experim. | | IPD |
| **Chalmers I** | 2013 | UK | Other | Case study | IPD |
| **Bergeris A** | 2018 | US | Survey | | IPD |
| **Tudur Smith C** | 2017 | UK | Other | Case study | Broader |
| **Vaduganathan M** | 2018 | US | Metrics | | IPD |
| **Merson L** | 2015 | Vietnam | Qualitative | | IPD |
| **Rowhani-Farid A** | 2016 | Australia | Survey | | IPD |
| **Rathi V** | 2012 | US | Survey | | IPD |
| **Ali J** | 2015 | US | Survey | | IPD |
| **Hopkins C** | 2016 | UK | Survey | | IPD |
| **Sydes M** | 2015 | UK | Metrics | Case study | IPD |
| **Polanin J** | 2019 | US | Experim. | | IPD |
| **Villain B** | 2015 | France | Survey | | IPD |
| **Asare A** | 2016 | US | Metrics | | IPD |
| **Strom B** | 2016 | US | Other | Metrics + survey | IPD |
| **Mello M** | 2005 | US | Survey | | IPD |
| **Rathi V** | 2014 | US | Survey | | IPD |
| **Huser V** | 2018 | US | Metrics | | IPD |
| **Chapman S** | 2014 | UK | Survey | | IPD |
| **Griswold M** | 2013 | US | Survey | | Broader |

| | | | | | |
|---|---|---|---|---|---|
| **Cheah PY** | 2015 | Thailand | Qualitative | | IPD |
| **Hee SW** | 2016 | UK | Other | Case study | IPD |
| **Geifman N** | 2015 | US | Metrics | | IPD |
| **Strom B** | 2014 | US | Metrics | | IPD |
| **Ross J** | 2016 | US | Survey | | IPD |
| **Boutron I** | 2016 | France | Survey | | Broader |
| **Vidal-Infer A** | 2018 | Spain | Survey | | Broader |
| **Krumholz H** | 2016 | US | Metrics | | IPD |
| **Tannenbaum S** | 2018 | US | Survey | | IPD |
| **Ross J** | 2017 | US | Metrics | | IPD |
| **Mello M** | 2018 | US | Survey | | IPD |
| **Chickramane A** | 2017 | India | Survey | | IPD |
| **Howe N** | 2018 | UK | Qualitative | | IPD |
| **Naudet F** | 2018 | US | Survey | | IPD |
| **Yuanyuan J** | 2017 | China | Survey | | IPD |
| **Spence O** | 2018 | US | Survey | | IPD |
| **Polanin J** | 2018 | US | Survey | | IPD |
| **Zhu C** | 2017 | US | Metrics | | IPD |
| **So D** | 2017 | Canada | Survey | | IPD |
| **Savage C** | 2009 | US | Survey | | IPD |
| **Kawahara T** | 2018 | Japan | Metrics | | IPD |
| **Goldacre B** | 2017 | UK | Survey | | IPD |
| **Pisani E** | 2017 | UK | Other | Metrics + Qualitative research | IPD |
| **Bertagnolli M** | 2017 | US | Metrics | | IPD |
| **Coady S** | 2017 | US | Metrics | | IPD |
| **Hopkins A** | 2018 | Australia | Survey | Survey | IPD |
| **Piwowar H** | 2007 | US | Survey | | IPD |
| **Laine C** | 2009 | US | Survey | | Broader |
| **Shmueli-Blumberg D** | 2013 | US | Metrics | | IPD |
| **de Vito N** | 2018 | UK | Survey | | Broader |

| | | | | | |
|---|---|---|---|---|---|
| **Nevitt S** | 2017 | UK | Survey | | IPD |
| **Ahmed I** | 2011 | UK | Survey | | IPD |
| **Navar A** | 2016 | USA | Metrics | | IPD |
| **Ebrahim S** | 2014 | USA | Survey | | IPD |
| **Vassar M** | 2020 | USA | Survey | 0 | IPD |
| **Cheah PY** | 2018 | Thailand | Qualitative | 0 | IPD |
| **Staham EE** | 2020 | USA | Survey | 0 | Broader |
| **Nutu D** | 2019 | Romania | Survey | 0 | IPD |
| **Aleixandre-Benavent R** | 2019 | Spain | Survey | 0 | IPD |
| **Ross JS** | 2018 | USA | Metrics | 0 | Broader |
| **Gorman DM** | 2019 | USA | Survey | 0 | IPD |
| **Bosserdt M** | 2019 | Germany | Survey | 0 | IPD |
| **Whitlock EP** | 2019 | USA | Survey | 0 | IPD |
| **Gabelica M** | 2019 | Croatia | Survey | 0 | IPD |
| **Gorman DM** | 2020 | USA | Survey | 0 | IPD |
| **Kaufmann I** | 2019 | UK | Survey | 0 | IPD |
| **Veroniki AA** | 2019 | Greece | Experim. | 0 | IPD |
| **Godolphin PJ** | 2019 | UK | Experim. | 0 | Broader |
| **Rowhani-Farid A** | 2020 | USA | Experim. | 0 | IPD |
| **Siebert M** | 2020 | France | Survey | 0 | IPD |
| **Mayer C** | 2019 | USA | Survey | 0 | IPD |
| **Gaba JF** | 2020 | France | Survey | 0 | Broader |
| **Colombo C** | 2019 | Italy | Survey | 0 | IPD |
| **Kochhar S** | 2019 | India | Metrics | 0 | IPD |
| **Broes S** | 2020 | Belgium | Qualitative | 0 | IPD |
| **Rollando P** | 2020 | France | Survey | 0 | Broader |
| **Schmidt H** | 2018 | Germany | Metrics | 0 | IPD |
| **Azar M** | 2020 | Canada | Survey | 0 | IPD |
| **Almaqrami BS** | 2020 | China | Survey | 0 | IPD |

| | | | | | |
|---|---|---|---|---|---|
| **Papageorgiou SN** | 2019 | Switzerland | Survey | 0 | iPD |
| **Miller J** | 2019 | USA | Survey | 0 | Broader |
| **Lovato L** | 2018 | USA | Metrics | 0 | IPD |
| **Kemper JM** | 2020 | Australia | Survey | 0 | IPD |
| **Johnson AL** | 2020 | USA | Survey | 0 | Broader |
| **Sherry C** | 2019 | USA | Survey | | Broader |
| **Pellen C** | 2020 | France | Survey | 0 | Broader |
| **Danchev V** | 2020 | USA | Survey | | IPD |
| **Li R** | 2020 | USA | Metrics | | IPD |

# Supplementary material 4: Published studies about YODA and CSDR

| Reference | Repository/ platform | No. of trials included in repository/ platform | No. of requests | No. of access to data | No. of publications | Assessment |
|---|---|---|---|---|---|---|
| Ross, 2017 | YODA | 189 | 73 | 50 | 2 | 6/2017 |
| Ross, 2018 | YODA | 270 | 100 | 82 | 11 | 8/2018 |
| Strom, 2016 | CSDR | 3049 | 177 | 144 | 1** | 5/2013-11/2015 |
| Schmidt, 2018 | CSDR | 3804 | 55 | 37 | 4 | 2004-2017 |
| Vadugan athan, 2018 | CSDR | 537*** | - | 30* | 3 | 1/2013-5/2017 |
| Kochar, 2019 | CSDR | >4000 | 473 | 222 | 41 | 2004-1/2019 |
| Navar, 2016 | CSDR, YODA, SOAR | >3000 | 234 | 113” | 1 | 2013-12/2015 |

*signed data use agreement
** 4 submitted for publication
*** cardiology trials

# Supplementary material 5: Published outputs from YODA (up to 1st July 2019) and CSDR (up to 31 August 2019)

| Published outputs | Title | Platform used | Identification of the proposal | Type of study | Request date |
|---|---|---|---|---|---|
| **Allott EH et al. 2017** | Statin Use, Serum Lipids, and Prostate Inflammation in Men with a Negative Prostate Biopsy: Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Smoking Is Associated with Acute and Chronic Prostatic Inflammation: Results from the REDUCE Study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Branche BL et al. 2017** | Sleep Problems are Associated with Development and Progression of Lower Urinary Tract Symptoms: Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Vidal AC et al. 2016** | Racial differences in prostate inflammation: results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2016** | Does Prostate Size Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2017** | Does Peak Urine Flow Rate Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Chronic baseline prostate inflammation is associated with lower tumor volume in men with prostate cancer on repeat biopsy: Results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Kent DM et al. 2016** | Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. | CSDR | 647 | Methodological | 29/10/2013 |
| **Baay M et al. 2017** | Background rates of disease in Latin American children from a rotavirus vaccine study. | CSDR | 651 | Secondary analysis | 11/03/2014 |
| **Le Noury J et al. 2015** | Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. | CSDR | 669 | Re-analysis | 27/01/2014 |
| **Nevitt SJ et al. 2017** | Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. | CSDR | 674 | Methodological | 15/05/2014 |
| **Nevitt SJ et al. 2017** | Antiepileptic drug monotherapy for epilepsy: a network meta-analysis of individual participant data. | CSDR | 674 | Meta-analysis | 15/05/2014 |

| | | | | | |
|---|---|---|---|---|---|
| **Forbess LJ et al. 2017** | Failure of a systemic lupus erythematosus response index developed from clinical trial data: lessons examined and learned. | CSDR | 911 | Secondary analysis | 25/07/2014 |
| **Dennis JM et al. 2018** | Evaluating associations between the benefits and risks of drug therapy in type 2 diabetes: a joint modeling approach. | CSDR | 930 | Secondary analysis | missing |
| **Dennis JM et al. 2018** | Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data. | CSDR | 930 | Secondary analysis | missing |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 945 | Secondary analysis | 23/04/2014 |
| **Mistry HB et al. 2017** | Model based analysis of the heterogeneity in the tumour size dynamics differentiates vemurafenib, dabrafenib and trametinib in metastatic melanoma. | CSDR | 946 | Secondary analysis | 28/05/2014 |
| **Muff S et al. 2018** | Bias away from the null due to miscounted outcomes? A case study on the TORCH trial. | CSDR | 977 | Re-analysis | 12/05/2014 |
| **Fragoso CAV et al. 2018** | Spirometric Criteria for Chronic Obstructive Pulmonary Disease in Clinical Trials of Pharmacotherapy. | CSDR | 993 | Secondary analysis | 28/02/2017 |
| **Devilliers H et al. 2016** | Minimal Clinically Important Differences for Generic Patient Reported Outcomes Tools in SLE | CSDR | 998 | Secondary analysis | missing |
| **Li-Kim-Moy J et al. 2018** | Impact of Fever and Antipyretic Use on Influenza Vaccine Immune Reponses in Children. | CSDR | 1000 | Secondary analysis | 08/09/2014 |
| **Blanco JR et al. 2017** | Impact of dolutegravir and efavirenz on immune recovery markers: results from a randomized clinical trial. | CSDR | 1028 | Secondary analysis | 23/09/2014 |
| **Borges NA et al. 2016** | Nonnucleoside Reverse-transcriptase Inhibitor- vs Ritonavir-boosted Protease Inhibitor-based Regimens for Initial Treatment of HIV Infection: A Systematic Review and Metaanalysis of Randomized Trials. | CSDR | 1058 | Meta-analysis | 18/08/2014 |
| **Dodd S et al. 2018** | Incidence and characteristics of the nocebo response from meta-analyses of the placebo arms of clinical trials of olanzapine for bipolar disorder. | CSDR | 1078 | Meta-analysis | 09/10/2014 |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 1079 | Secondary analysis | 12/10/2014 |
| **Emamikia S et al. 2017** | Relationship between glucocorticoid dose and adverse events in systemic lupus erythematosus: data from a randomized clinical trial. | CSDR | 1084 | Secondary analysis | 20/02/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Gruber JF et al. 2018** | Timing and predictors of severe rotavirus gastroenteritis among unvaccinated infants in low- and middle-income countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Gruber JF et al. 2018** | Timing of Rotavirus Vaccine Doses and Severe Rotavirus Gastroenteritis Among Vaccinated Infants in Low- and Middle-income Countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Schwartz LM et al. 2016** | Rotavirus vaccine effectiveness in low-income settings: An evaluation of the test-negative design. | CSDR | 1090 | Secondary analysis | 15/04/2015 |
| **Hilkens NA et al. 2016** | Blood pressure levels and the risk of intracerebral hemorrhage after ischemic stroke. | CSDR | 1100 | Secondary analysis | 13/01/2015 |
| **Hieronymus F et al. 2017** | Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. | CSDR | 1103 | Meta-analysis | missing |
| **Waljee AK et al. 2018** | Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. | CSDR | 1136 | Secondary analysis | 13/08/2015 |
| **Hadjichrysanthou C et al. 2016** | Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. | CSDR | 1137 | Secondary analysis | 16/04/2015 |
| **Voysey M et al. 2017** | The Influence of Maternally Derived Antibody and Infant Age at Vaccination on Infant Vaccine Responses : An Individual Participant Meta-analysis. | CSDR | 1141 | Meta-analysis | 22/07/2015 |
| **Radua J et al. 2017** | Meta-Analysis of the Risk of Subsequent Mood Episodes in Bipolar Disorder. | CSDR | 1148 | Meta-analysis | 30/01/2015 |
| **de Vries YA et al. 2018** | Initial severity and antidepressant efficacy for anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: An individual patient data meta-analysis. | CSDR | 1173 | Meta-analysis | 30/06/2015 |
| **Zafack JG et al. 2019** | Adverse events following immunisation with four-component meningococcal serogroup B vaccine (4CMenB): interaction with co-administration of routine infant vaccines and risk of recurrence in European randomised controlled trials. | CSDR | 1224 | Secondary analysis | missing |
| **Sturm A et al. 2017** | Evaluating the Hierarchical Structure of ADHD Symptoms and Invariance Across Age and Gender. | CSDR | 1292 | Methodological | 29/07/2015 |
| **Oon S et al. 2019** | Lupus Low Disease Activity State (LLDAS) discriminates responders in the BLISS-52 and BLISS-76 phase III trials of belimumab in systemic lupus erythematosus. | CSDR | 1320 | Secondary analysis | missing |
| **Craig K et al. 2017** | More of what works: Detection of informative sites during the conduct of clinical trials using machine learning | CSDR | 1323 | Methodological | 21/10/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Bauza C et al. 2018** | Determining the Joint Effect of Obesity and Diabetes on All-Cause Mortality and Cardiovascular-Related Mortality following an Ischemic Stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Bauza C et al. 2018** | Determining the joint effect of obesity and diabetes on functional disability at 3-months and on all-cause mortality at 1-year following an ischemic stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Tajgardoon M et al. 2018** | A Novel Representation of Vaccine Efficacy Trial Datasets for Use in Computer Simulation of Vaccination Policy. | CSDR | 1374 | Secondary analysis | 25/05/2016 |
| **Berenguer J et al. 2019** | Mathematical modeling of HIV-1 transmission risk from condomless anal intercourse in HIV-infected MSM by the type of initial ART. | CSDR | 1403 | Secondary analysis | missing |
| **Hilkens NA et al. 2017** | Predicting Major Bleeding in Ischemic Stroke Patients With Atrial Fibrillation. | CSDR | 1455 | Secondary analysis | 03/06/2016 |
| **Kerr SJ et al. 2017** | The FDA snapshot algorithm may overestimate the efficacy of initial art | CSDR | 1456 | Methodological | missing |
| **Samara MT et al. 2017** | Initial symptom severity of bipolar I disorder and the efficacy of olanzapine: a meta-analysis of individual participant data from five placebo-controlled studies. | CSDR | 1457 | Meta-analysis | 08/06/2016 |
| **Hopkins AM et al. 2018** | Risk Factors for Severe Diarrhea with an Afatinib Treatment of Non-Small Cell Lung Cancer: A Pooled Analysis of Clinical Trials. | CSDR | 1475 | Meta-analysis | missing |
| **Peters EM et al. 2018** | Melancholic Symptoms in Bipolar II Depression and Responsiveness to Lamotrigine in an Exploratory Pilot Study. | CSDR | 1569 | Secondary analysis | 01/11/2016 |
| **de Vries YA et al. 2018** | Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: individual patient data meta-analysis. | CSDR | 1575 | Meta-analysis | 11/10/2016 |
| **Gemeinsamer Bundesausschuss, 2019** | Nutzenbewertungsverfahren zum Wirkstoff Sitagliptin | CSDR | 1593 | Secondary analysis | 04/11/2016 |
| **Hopkins AM et al. 2019** | Effect of early adverse events on response and survival outcomes of advanced melanoma patients treated with vemurafenib or vemurafenib plus cobimetinib: A pooled analysis of clinical trial data. | CSDR | 1599 | Meta-analysis | missing |
| **Carbon M et al. 2018** | Tardive dyskinesia risk with first- and second-generation antipsychotics in comparative randomized controlled trials: a meta-analysis. | CSDR | 1624 | Meta-analysis | missing |

| | | | | | |
|---|---|---|---|---|---|
| **Schwarzman LS et al. 2018** | The Association of Previous Prostate Biopsy Related Complications and the Type of Complication with Patient Compliance with Rebiopsy Scheme. | CSDR | 1626 | Secondary analysis | 16/12/2016 |
| **Voysey M et al. 2017** | Prevalence and decay of maternal pneumococcal and meningococcal antibodies: A meta-analysis of type-specific decay rates. | CSDR | 1629 | Meta-analysis | 26/09/2016 |
| **Shapiro W et al. 2018** | Salmeterol Combined with Fluticasone Reduces Exacerbations More Effectively in Chronic Bronchitis Associated with Chronic Obstructive Pulmonary Disease: A Post-hoc Analysis of the TORCH Trial | CSDR | 1640 | Secondary analysis | 28/02/2017 |
| **Parodis I et al. 2018** | Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Established organ damage reduces belimumab efficacy in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Predictors of low disease activity and clinical remission following belimumab treatment in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hernández-Breijo B et al. 2019** | Antimalarial agents diminish while methotrexate, azathioprine and mycophenolic acid increase BAFF levels in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hopkins AM et al. 2019** | Predictors of Long-Term Disease Control and Survival for HER2-Positive Advanced Breast Cancer Patients Treated With Pertuzumab, Trastuzumab, and Docetaxel. | CSDR | 1741 | Secondary analysis | missing |
| **Janciauskiene S et al. 2019** | Serum Levels of Alpha1-antitrypsin and Their Relationship With COPD in the General Spanish Population. | CSDR | 2084 | Secondary analysis | missing |
| **Storgaard H et al. 2016** | Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis. | YODA | 2014-0340 | Meta-analysis | 19/11/2014 |
| **Christian KE et al. 2019** | Gender Differences and Other Factors Associated with Weight Gain Following Initiation of Infliximab: A Post Hoc Analysis of Clinical Trials. | YODA | 2014-0334 | Meta-analysis | 26/11/2014 |
| **Wang R et al. 2018** | Comparative Efficacy of Tumor Necrosis Factor-α Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis. | YODA | 2014-0291 | Meta-analysis | 08/12/2014 |
| **Waljee AK et al. 2017** | External Validation of a Thiopurine Monitoring Algorithm on the SONIC Clinical Trial Dataset. | YODA | 2014-0401 | Secondary analysis | 20/01/2015 |
| **Mospan GA et al. 2017** | 5-Day versus 10-Day Course of Fluoroquinolones in Outpatient Males with a Urinary Tract Infection (UTI). | YODA | 2015-0514 | Secondary analysis | 26/05/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Singh S et al. 2018** | Impact of Obesity on Short- and Intermediate-Term Outcomes in Inflammatory Bowel Diseases: Pooled Analysis of Placebo Arms of Infliximab Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Singh S et al. 2018** | Obesity and Response to Infliximab in Patients with Inflammatory Bowel Diseases: Pooled Analysis of Individual Participant Data from Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Spertus J et al. 2018** | Risk of weight gain for specific antipsychotic drugs: a meta-analysis. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Spertus J et al. 2019** | Bayesian Meta-analysis of Multiple Continuous Treatments with Individual Participant-Level Data: An Application to Antipsychotic Drugs. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Zou X et al. 2018** | The role of PANSS symptoms and adverse events in explaining the effects of paliperidone on social functioning: a causal mediation analysis approach. | YODA | 2016-0716 | Secondary analysis | 24/02/2016 |
| **World Health Organization 2017** | WHO report (appendix) | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Mbuagbaw L et al. 2019** | Outcomes of Bedaquiline Treatment in Patients with Multidrug-Resistant Tuberculosis. | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Corbett M et al. 2017** | Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. | YODA | 2016-0897 | Meta-analysis | 19/05/2016 |
| **Gay HC et al. 2017** | Feasibility, Process, and Outcomes of Cardiovascular Clinical Trial Data Sharing: A Reproduction Analysis of the SMART-AF Trial. | YODA | 2016-0912 | Re-analysis | 07/06/2016 |
| **Schneider-Thoma J et al. 2018** | Second-generation antipsychotic drugs and short-term mortality: a systematic review and meta-analysis of placebo-controlled randomised controlled trials. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Schneider-Thoma J et al. 2019** | Second-generation antipsychotic drugs and short-term somatic serious adverse events: a systematic review and meta-analysis. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Teply BA et al. 2019** | Risk of development of visceral metastases subsequent to abiraterone vs placebo: An analysis of mode of radiographic progression in COU-AA-302. | YODA | 2016-1057 | Secondary analysis | 01/09/2016 |
| **Loubersac T et al. 2019** | Neutrophil-to-lymphocyte Ratio as a Predictive Marker of Response to Abiraterone Acetate: A Retrospective Analysis of the COU302 Study. | YODA | 2016-1103 | Secondary analysis | 23/11/2016 |
| **Martin LJ et al. 2018** | Identification of subgroups of metastatic castrate-resistant prostate cancer (mCRPC) patients treated with abiraterone | YODA | 2016-1122 | Secondary analysis | 23/11/2016 |

| | | | | | |
|---|---|---|---|---|---|
| | plus prednisone at low- vs. high-risk of radiographic progression: An analysis of COU-AA-302. | | | | |
| **Waljee AK et al. 2019** | Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. | YODA | 2016-1176 | Secondary analysis | 01/03/2017 |
| **Kubo K et al. 2018** | Placebo effects in adult and adolescent patients with schizophrenia: combined analysis of nine RCTs. | YODA | 2017-1676 | Meta-analysis | 24/05/2017 |
| **Kumagai F et al. 2018** | Early Placebo Improvement Is a Marker for Subsequent Placebo Response in Long-Acting Injectable Antipsychotic Trials for Schizophrenia: Combined Analysis of 4 RCTs. | YODA | 2017-1701 | Meta-analysis | 01/06/2017 |
| **Yiu ZZN et al. 2019** | A standardization approach to compare treatment safety and effectiveness outcomes between clinical trials and real-world populations in psoriasis. | YODA | 2017-1706 | Methodological | 08/08/2017 |
| **Narula N et al. 2018** | Patient-Reported Outcomes and Endoscopic Appearance of Ulcerative Colitis: A Systematic Review and Meta-analysis. | YODA | 2017-2031 | Meta-analysis | 30/08/2017 |
| **Singh S et al. 2018** | No Benefit of Concomitant 5-Aminosalicylates in Patients With Ulcerative Colitis Escalated to Biologic Therapy: Pooled Analysis of Individual Participant Data From Clinical Trials. | YODA | 2017-2306 | Meta-analysis | 25/09/2017 |
| **Singh S et al. 2019** | Efficacy and Speed of Induction of Remission of Infliximab vs Golimumab for Patients With Ulcerative Colitis, Based on Data From Clinical Trials. | YODA | 2018-3121 | Meta-analysis | 21/05/2018 |

# Status, use and impact of sharing Individual Participant data from clinical trials: a scoping review

C. Ohmann et al.

**Scoping Reviews (PRISMA.ScR) Checklist:**

| Section | Item | Covered |
|---|---|---|
| Title | Title | yes |
| Abstract | Structured summary | yes |
| Introduction | Rationale | yes |
| | Objectives | yes |
| Methods | Protocol and registration | yes |
| | Eligibility criteria | yes |
| | Information sources | yes |
| | Search | yes |
| | Selection of sources of evidence | yes |
| | Data charting process | yes |
| | Data items | yes |
| | Critical appraisal of individual sources of evidence | yes |
| | Synthesis of results | yes |
| Results | Selection of sources of evidence | yes |
| | Characteristics of sources of evidence | yes |
| | Critical appraisal within sources of evidence | yes |
| | Results of individual sources of evidence | yes |
| | Synthesis of results | yes |
| Discussion | Summary of evidence | yes |
| | Limitations | yes |
| | Conclusions | yes |
| Funding | Funding | yes |

BMJ Open

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-049228.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 01-Jul-2021 |
| Complete List of Authors: | Ohmann, Christian; European Clinical Research Infrastructure Network (ECRIN),<br>Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre<br>Siebert, Maximilian; University Rennes, CHU Rennes, CIC 1414 (Centre d'Investigation Clinique de Rennes)<br>Motschall, Edith ; University of Freiburg Faculty of Medicine, Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg,<br>Naudet, Florian; University Rennes, CHU Rennes, INSERM CIC 1414 (Centre d'Investigation Clinique de Rennes) |
| Primary Subject Heading: | Health informatics |
| Secondary Subject Heading: | Health informatics |
| Keywords: | Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Information management < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review

**(Date: 01.07.2021)**

**Authors**

C. Ohmann, ORCID: 0000-0002-5919-1003, christianohmann@outlook.de
D. Moher, ORCID: 0000-0003-2434-4206, dmoher@ohri.ca
M. Siebert, ORCID: 0000-0003-4385-5773, maximiliansiebert91@gmail.com
E. Motschall, ORCID: 0000-0001-7795-242X, motschall@imbi.uni-freiburg.de
F. Naudet, ORCID: 0000-0003-3760-3801, floriannaudet@gmail.com

**Corresponding author:**

Prof. Dr. Christian Ohmann
European Clinical Research Infrastructure Network (ECRIN)
Kaiserswerther Str. 70
40477 Düsseldorf
Germany
Tel.: +49 211 5142490
Mobile: + 49 16097945385
Email: christianohmann@outlook.de

**Co-authors :**

David Moher
Ottawa Hospital Research Institute
Ottawa Methods Centre
Ottawa, On
Canada

Maximilian Siebert, Florian Naudet
Université de Rennes 1
CHU Rennes
INSERM, CIC 1414 (Centre d'Investigation Clinique de Rennes)
Rennes, Bretagne
France

Edith Motschall
Institute of Medical Biometry and Statistics
Faculty of Medicine and Medical Center
University of Freiburg
Germany

**Key words:**

clinical trial, individual participant data, data sharing, scoping review, impact

**Word count:**

12316

**ABSTRACT**

**Objectives**
To explore the impact of data-sharing initiatives on the intent to share data, on actual data-sharing, on the use of shared data and on research output and impact of shared data.
**Eligibility criteria**
All studies investigating data-sharing practices for individual participant data (IPD) from clinical trials.
**Sources of evidence**
We searched the Medline database, the Cochrane Library, the Science Citation Index Expanded and the Social Sciences Citation Index via Web of Science, and preprints and proceedings of the International Congress on Peer Review and Scientific Publication. In addition, we inspected major clinical trial data-sharing platforms, contacted major journals/publishers, editorial groups and some funders.
**Charting methods**
Two reviewers independently extracted information on methods and results from resources identified using a standardised questionnaire. A map of the extracted data was constructed and accompanied by a narrative summary for each outcome domain.
**Results**
93 studies identified in the literature search (published between 2001-2020, median: 2018) and 5 from additional information sources were included in the scoping review. Most studies were descriptive and focused on early phases of the data-sharing process. While the willingness to share IPD from clinical trials is extremely high, actual data-sharing rates are suboptimal. A survey of journal data suggests poor to moderate enforcement of the policies by publishers. Metrics provided by platforms suggest that a large majority of data remains unrequested. When requested, the purpose of the re-use is more often secondary analyses and meta-analyses, rarely re-analyses. Finally, studies focused on the real impact of data-sharing were rare and used surrogates such as citation metrics.
**Conclusions**
There is currently a gap in the evidence base for the impact of IPD sharing, which entails uncertainties in the implementation of current data-sharing policies. High level evidence is needed to assess whether the value of medical research increases with data-sharing practices.

**Strengths and limitations of this study**

- Exhaustive review of both the literature and the main initiatives in data-sharing

- Analysis of the full data-sharing process covering intention to share, actual sharing, use of shared data, research output and impact

- Retrieval and synthesis of information proved to be difficult because of a very siloed landscape where each initiative/platform operates with its own metrics

- Data-sharing is a moving target in a rapidly changing environment with more and more new initiatives.

- Only a limited research output from data-sharing is available so far

**Funding**

No specific funding for this review.
FN's work on data-sharing is supported by a grant from the French National Research Agency – ANR (Reproducibility in Therapeutic Research / ReITheR: Agence Nationale de la Recherche, ANR-17-CE36-0010-01).
CO's work on data-sharing is supported by funding from the European Union's Horizon 2020 Research and Innovation Programme (CORBEL, under grant agreement n° 654248).
DM is supported by an Ottawa University Research Chair (grant number: N/A).

**Competing interests**

None of the authors have any competing interests.

**Author's contribution**

CO, DM and FN developed the study protocol. The search strategy was developed and implemented by EM. The selection of sources of evidence and the assessment was performed by CO and FN. Contact with initiatives/platforms/journals/publishers was performed by MS. In case of disagreements, these were resolved by consensus and, when necessary, in consultation with DM. The first draft of the manuscript was written by CO and FN. All authors have revised and approved the final manuscript.

**Data sharing statement**

All data relevant to the study is included in the article or uploaded as supplementary Information.

**Acknowledgements**

We thank Angela Swaine Verdier for revising the English in the manuscript.

## INTRODUCTION

### Rationale

Data sharing is increasingly recognized as a key requirement in clinical research.[1] In any discussion about clinical trial data-sharing the emphasis is naturally on the data sets themselves, but data-sharing is much broader. Besides the individual participant data sets, other clinical trial data sources should be made available for sharing (e.g., protocols, clinical study reports, statistical analysis plans, blank consent forms) to enable a full understanding of any data set. In this scoping review, there is a focus on the sharing of individual participant data from clinical trials.

Within clinical research, data-sharing can enhance reproducibility and the generation of new knowledge, but it also has an ethical and economic dimension.[2] Scientifically, sharing makes it possible to compare or combine the data from different studies, and to more easily aggregate it for meta-analysis. It enables conclusions to be re-examined and verified or, occasionally, corrected, and it can enable new hypotheses to be tested. Sharing can therefore increase data validity, but it also draws more value from the original research investment, as well as helping to avoid unnecessary repetition of studies. Agencies and funders are referring more and more to the economic advantages of data reuse. Ethically, data-sharing provides a better way to honour the generosity of clinical trial participants, because it increases the utility of the data they provide. Despite the high potential for sharing clinical trial data, the launch and implementation of several data-sharing initiatives and platforms, and outstanding examples related to the value of data-sharing,[3] to date data-sharing is not the norm in clinical research, unlike many other scientific disciplines.[4] One major hurdle is that clinical trial data concerns individuals and their health status, and as such requires specific measures to protect privacy.

To support sharing of individual participant data (IPD) in clinical trials, several organisations have developed generic principles, guidance and practical recommendations for implementation. In 2016, the International Committee of Medical Journal Editors (ICMJE), a small group of medical journal editors, published an editorial[5] stating that "it is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk". The ICMJE considers that there is an implicit social contract imposing an ethical obligation for the results to lead to the greatest possible benefit to society. The ICMJE proposed to require that de-identified IPD is made publicly available no later than 6 months after publication of the main trial results. This time lapse would be useless for public health emergencies like COVID-19. However, the ICMJE proposal triggered debate, and a large number of trialists were reluctant to adopt this new norm[6] on account of the feasibility of the proposed requirements, the resources required, the real or perceived risks to trial participants, and the need to protect the interests of patients and researchers.[7]

Despite the cultural shift towards sharing clinical trial data and the major commitment of scientific organisations, funders and initiatives, overall there is still a lack of effective policies in the biomedical literature to ensure that underlying data is maximally available and reusable. The only requirement appears to be a data management plan or a data-sharing plan. A few journals require data-sharing and, for those who do require data-sharing, guidelines are heterogeneous and somewhat ambiguous.[8] Nevertheless, some innovative and progressive funders (e.g. Wellcome Trust, Bill & Melinda Gates Foundation), and publishers/journals (e.g. Public Library of Science (PLOS) [in 2014], The British Medical Journal (BMJ)) [2009-2015], have adopted strong data-sharing policies. As part of a wider cultural shift towards more open science, there have been various attempts to explore how clinical researchers can best plan for data-sharing and prepare their 'raw' IPD so that it becomes available to others[9] – albeit often under controlled access conditions rather than simply being publicly available on-line[10] - and can structure that data to make it FAIR (findable, accessible, interoperable and reusable).[11] Meanwhile several data-sharing platforms and repositories are available and in use to provide practical support for the data-sharing process in clinical research (e.g. Yale University Open Data Access (YODA) launched [in 2011], ClinicalStudyDataRequest.com (CSDR) [launched in 2013], Vivli [launched in 2018]. A considerable number of individual studies have been performed to access and explore the sharing of data from clinical trials under different circumstances and within different frameworks. What is strongly needed is a scoping review providing an overview of the status of implementation of data-sharing as a whole and the implications originating from the available evidence.

### Objectives

In this scoping review we explored the impact of data-sharing initiatives on the willingness to share data, the status of data-sharing, the use of shared data and the impact of research outputs from shared data.

## METHODS

### Protocol and registration

The study protocol was registered on the Open Science Framework on September the 12th 2018 (registration number: osf.io/pb8cj). The protocol followed the methodology manual published by the Joanna Briggs Institute for scoping reviews.[12] Methods and results are reported using the PRISMA (Preferred Reporting Items for systematic Reviews and Meta Analyses) extension for scoping reviews (PRISMA-ScR).[13]

### Eligibility criteria

The following eligibility criteria for studies were used:

All study designs were eligible, including case studies, surveys, metrics and experimental studies, using qualitative or quantitative methods. Only published or unpublished reports (e.g. pre-prints, congress presentations, non-indexed information such as websites) in English, German, French or Spanish were considered.

We included all studies and reports 1/ providing information on current IPD data-sharing practices for clinical trials and 2/ reporting on one or more of five outcome domains defined according to the data-sharing process presented in **Box 1**.

**1. Intention to share data**

There is an intention to share data, expressed by a stakeholder (e.g., sponsor/PI, funder). This can be done by a written data-sharing commitment or by a declaration included in the trial registration. This also includes surveys on attitudes towards data-sharing.

**2. Actual data-sharing**

Data is truly made available for data-sharing to secondary users. This is important because there are cases known where the data is offered for sharing but sharing does not take place, as a result of a possible hidden agenda or change in plans.

**3. Use of shared data**

Shared data can be used for various purposes. It can be used as background for research, usually not leading to research outputs. This covers use for education, researcher training and understanding of data. Study types that should lead to new research outputs include 1/ validation/reproducibility of results, 2/ further additional analyses (prognostic models, decision-support, subgroup analyses, etc.) and 3/ IPD meta-analyses.

**4. Research outputs from shared data**

Research outputs are scientific presentations, reports and publications.

**5. Impact of research output from shared data**

Research output from shared data can have an impact on medical research (e.g. development of new hypotheses and methods) and/or medical health (e.g. changes in treatment via guidelines).

**Box 1:** Definitions used for the 5 outcome domains

In the scoping review only data-sharing of IPD from clinical trials was considered. We defined clinical trials following the clinicaltrials.gov definition: "a clinical study is a research study involving human volunteers (also called participants) that is intended to add to medical knowledge. There are two types of clinical studies: interventional studies (also called clinical trials) and observational studies. Clinical trial is another name for an interventional study."[14] We therefore considered any interventional clinical studies (no matter whether they were randomised), and we did not consider studies on data-sharing concerning observational and non-clinical studies (e.g. on genomics) nor different fields outside medicine (e.g. economics).
We included studies that investigated and reported information on current data-sharing practices performed without restrictions in terms of promotional initiatives, type of repository or platform (see **Box 2** for definitions) and that promoted data-sharing practices (e.g. at editorial level, at funder level, at research level etc.). We considered many different types of studies (e.g. experimental studies, surveys, metrics, quality assurance studies, qualitative research, reviews, reports), as the inclusion criteria were not method-specific but rather content-specific**.**

**Initiatives**

Major activities of an organization (or a network of several organizations) to actively promote data-sharing in this area (e.g. Pharmaceutical Research and Manufacturers of America (PHRMA)/European Federation of Pharmaceutical Industries and Associations (EFPIA), Nordic Trial Alliance, Institute of Medicine (IOM), ICMJE, Research Data Alliance (RDA)).

**Repository**

Large database infrastructures set up to manage, share, access and archive researchers' datasets from clinical trials. Repositories can be specialised and dedicated to specific disciplines (e.g. FreeBird, Biological Specimen and Data Repository Information Coordination Center (BioLINCC) or more general (e.g. FigShare, Dryad).

**Platform**

A computer environment where researchers can find datasets from clinical trials across different repositories, and where additional functionalities (e.g. protected analysis environment) are provided (e.g. CSDR, YODA, project Data Sphere, Github).

**Box 2:** Definitions used for initiatives, repository and platform

**Information sources**

The identification of studies was performed in two complementary stages:

a) A systematic literature search in bibliographic databases (MEDLINE databases, Cochrane Library, Science Citation Index Expanded and Social Science Citation Index). In addition, preprint servers and proceedings were searched
b) Inspection of and if required contacts with known information sources (e.g. webpages, documents and reports from platforms, funder, publisher) to explore whether they had an evaluation component and provided detailed research output from shared data (see supplementary material 1).

Between 25/01/2019 and 12/06/2019 (with an update on 02/11/2020), one researcher (MS) inspected (and when necessary contacted) major clinical trial data-sharing platforms to explore whether they had an evaluation component and provided details of research output from shared data (see **Supplementary Material 1**). Similarly, in the same time period, the researcher contacted major journals and/or publishers and/or editorial groups (The BMJ, PLOS, The Annals of Internal Medicine, BioMedCentral (Springer/Nature), Faculty of 1000 Research (F1000Research)). These journals/publishers were targeted because they had either an early or a robust data-sharing policy (NEJM, Lancet and JAMA had no data-sharing policy before the 2018 ICMJE policy). Some funders (see **Supplementary Material 1**) were also contacted, and preprints repositories were explored (bioRxiv, PeerJ, Preprints.org, PsyArXiv and MedRxiv. For the sake of completeness, ASAPbio

(Accelerating Science and Publication in biology) and the Center for Open Science were also contacted for the same information, as well as three International Congress on Peer Review and Scientific Publication conference abstracts. In addition, when relevant references were found in various papers these references were included (snowballing searches).

**Search**

On 29/10/2018 (update on 12/09/2020), one researcher (EM) searched the Medline databases for indexed and non-indexed citations via Ovid from Wolters Kluwer, the Cochrane Library via Wiley, Science Citation Index Expanded and Social Sciences Citation Index via Web of Science from Clarivate Analytics for articles meeting our inclusion criteria.

The detailed search terms for the MEDLINE databases, the Cochrane Library and the Web of Science databases can be found in **Supplementary Material 2**. The main search strategy developed by CO, DM und FN was peer-reviewed independently (by a senior medical documentalist, EM who joined the team subsequently) using evidence-based guidelines for Peer Review of Electronic Search Strategies (PRESS).[15] Discrepancies were resolved between the authors, and EM performed the search. All references were managed and de-duplicated using a reference manager system (Endnote).

On 23/01/2019 (update on 02/11/2020), two researchers (MS and FN) independently searched for relevant pre-prints on OSF PREPRINTS using the search function to find all papers relevant to medicine with the following keyword (trial* OR random*). On 29/01/2019, the two researchers independently searched the proceedings of the three latest International Congress on Peer Review and Scientific Publication reports for relevant abstracts (2009, 2013 and 2017).

**Selection of sources of evidence**

The selection of sources of evidence was performed by two independent reviewers (CO and FN). Contact with initiatives/platforms/journals/publishers was made by a single reviewer (MS). In case of disagreements, these were resolved by consensus between CO and FN and, when necessary, in consultation with a third reviewer (DM).

**Data charting process**

We developed a data collection form and pilot-tested it on 10 randomly selected research papers which were later included in our final study. In case of disagreement, these were resolved by consensus and, when necessary, in consultation with a third reviewer (DM).

**Data items**

For each research paper included according to the selection criteria we extracted: 1/ basic information on the paper (type of study exploring data-sharing practices, authors, year, references, and type of initiative and/or repository and/or platform studied), 2/ information on the material shared (sharing of data, code, programs and material), 3/ whether it reported data about one or more of the five outcomes domains defined box 1, 4/ how these outcome domains were assessed, and 5/ a qualitative description of the main results observed on these outcomes.

For each data-sharing platform, publisher and funder providing detailed research output from shared data, we extracted the following information (authors, date of request, date of publication, type of re-use). We initially planned to describe the scale of re-use in qualitative terms and the observed results of the re-use (i.e. “positive” or “negative” study) but these two characteristics were difficult to extract with very poor inter-rater agreement and we decided not to detail them.

**Critical appraisal of individual sources of evidence**

The studies included were classified according to study type (e.g. survey, metrics, experimental). Potentially relevant characteristics of studies included with regard to their internal-external validity and risk of bias were

not assessed systematically with a specific tool, but explored when one of the two reviewers considered it relevant, and in this case each study was thoroughly discussed between the reviewers.

**Synthesis of results**

No outcome was prioritized since there was no quantitative synthesis for this study. All outcomes were described separately in sections corresponding to the outcome domain and subsections corresponding to similar types of initiative. Our plan for the presentation of results was specified in our protocol and organized into 1/ different sections corresponding to the key concepts detailed in the data-sharing pipeline (intention to share data, actual data-sharing, results of re-use, output from data-sharing, impact of data-sharing) and 2/ different subsections corresponding to the different contexts and actors involved in the data-sharing pipeline (e.g. targeted group for intention to share data or type of use for re-use of shared data)). A summary of the data extracted from the papers included was constructed in tabular form with basic characteristics, and was accompanied by a narrative summary describing all results observed in the light of the review objective and *question/s.* Usually, individual studies were summarized in a short text with descriptive statistics of the main results (numbers, percentages), when appropriate visual representations of the data extracted were provided.

**Patient and public involvement**

There was no patient or public involvement in this scoping review.

**Changes to the initial protocol**

We initially planned to contact leading authors in the field to ask whether they were aware of other unpublished initiatives, but this was not done as it was difficult to identify relevant authors. We found relevant references about data-sharing policies including both clinical trials and observational studies, without making a distinction. These references were included in the scoping review and this point was discussed in the text.

**RESULTS**

**Selection of sources of evidence**

A total of 3024 records were identified, 3,005 records (1991 + 1014 in the update) were retrieved by database search (2141 without duplicates). An additional 8 records were identified by screening the proceedings of the last three International Congress on Peer Review and Scientific Publication conference abstracts and ten records by snowballing searches. One additional relevant record was identified after screening 630 identified pre-prints. We screened all irrelevant records by title and abstract, leaving 409 possibly relevant references which were eligible for full-text screening. Subsequently, 316 references were excluded, leaving 93 reports that met the inclusion criteria (**Figure 1**). We inspected websites and when needed contacted 48 initiatives/platforms/journals (we actually screened 49 but Supporting Open Access for Research Initiative (SOAR) is now integrated into Vivli): 23 data-sharing platforms, 13 funding organisation, 5 journals, 5 pre-print repositories and 2 other initiatives. For 33 of these different sources, there was no evaluation component and for 10 additional contacts we received no answer as to whether they had an evaluation component and/or any data. 4 data-sharing platforms (CSDR, YODA, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Vivli) and 1 funding organisation (Medical Research Council United Kingdom (MRC UK)) provided some additional data (online metrics and or data about its policy) (**Figure 1**) which was extracted in June 2019 and updated in December 2020.

**Characteristics of sources of evidence**

Of the 93 reports, 5 were classified as experimental studies, 58 as surveys, 19 as metrics, 5 as qualitative research and 6 as other (4 case studies, 1 metrics & survey, 1 metrics and qualitative). The median year of publication was 2018 (range [2001-2020]). The vast majority of these studies were from North America (50,

54%), Europe (16, 17 %) and the UK (15, 16%). Eight (9%) were from Asia and 4 (4 %) from Australia. Most (78, 84 %) were focused on IPD data-sharing while the remaining 15 (16 %) adopted a wider definition of the material shared (e.g. by including protocols, codes). Thirty-eight reports (41 %) were focused on data-sharing in publications/journals, 23 (25 %) on data repositories, 8 (9 %) on data-sharing by various institutions, 4 (4 %) on trial registries and 20 (21 %) in various other contexts (see **Supplementary Material 3** which presents study characteristics in detail).

**Collating and summarising the data**

**Figure 2** shows the proportion of the 93 references exploring each outcome domain. In an effort to create a useful synthesis of results, we collated results on each outcome from each publication and organised them into the pre-specified categories. **Figure 3** presents a detailed overview of the different outcome domains and the related outcomes used in the 93 different references included, organised by type of research.

**Critical appraisal of sources of evidence**

In general, there was a high risk of bias, especially due to study design (e.g. surveys with low response rates and absence of experimental design). As stated in the methods, this was not assessed systematically. If available, we have tried to present this information in the narrative part of the review.

**Results for individual sources of evidence: intentions to share data**

***Clinical Trialists***

***Surveys of attitudes***

Four surveys investigating intention to share data by trialists reported high data-sharing rates of around 75% or more (see **Figure 4**). These surveys targeted authors of published trials and in one study reviewers in a Cochrane group (where the majority of respondents had been involved in a randomised controlled trial (RCT)). The studies differed by different estimations of data-sharing rates, different selection criteria and/or survey methods. Response rates were comparable across the surveys (42-58%). Reviewers in the Cochrane IPD meta-analysis group were strongly in favour of a central repository and of providing IPD for central storage (83%)[20]. In the survey by Rathi et al.[16], 74% and 72% respectively thought that sharing de-identified data through data repositories should be required and that investigators should be required to share de-identified data in response to individual requests. However, only 18% indicated that they were required by the trial funder to place the trial data in a repository. In this survey, support for data-sharing did not differ on trialist or trial characteristics.[17] Trialists in Western Europe indicated they had shared or would share data in order to receive academic benefits or recognition more frequently than those from the USA or Canada (58 versus 31%). The most academically productive trialists less frequently indicated they had withheld or would withhold data in order to protect research subjects (24 versus 40% for the least productive), as did those who had received industry funding compared to those who had not (24 versus 43%). The survey by Tannenbaum, 2018[18] suggested that willingness to share data could depend on the intended re-use of the data (97% of respondents were willing to share data for a meta-analysis versus 73% for a re-analysis). For secondary analyses, the willingness to share was largely influenced by respondents' willingness to conduct a similar analysis. In addition, willingness to share was more marked after 1 year than after 6 months. In the fourth survey on trials published in Chinese medical journals, the overwhelming majority (87%) stated that they endorsed data-sharing.[19]

*Metrics of data-sharing statements in journal articles*

Intentions to share data for trialists were less clear for data-sharing statements in published journal articles (although this section is not specific to clinical trials) (see **Figure 4**). Depending on the journals considered, the rates vary from less than 5 % to around 25%. An analysis of the first year after the Annals of Internal Medicine policies encouraged data-sharing[20] found that data was available without condition for 4%, with conditions for 57%, and unavailable for 38%. Over the first 4 years data was available without condition for 7%, with

conditions for 47%, and unavailable for 46% of research articles. 9% and 22 % of 160 randomly sampled research articles in the BMJ from 2009 to 2015 made data available or indicated the availability of their data sets.[21] Among 60 randomized cardiovascular interventional trials registered on ClinicalTrials.gov[22] up to 2015 with >5000 enrollment, sponsored by one of the top 20 pharmaceutical companies in terms of 2014 global sales, IPD was available for 15 trials (25%) amounting to 204 452 patients, unavailable for 15 trials (25%) and undetermined for the remaining 50 %, because of either no response or requirements for a full proposal. Reasons for non-availability were: co-sponsor did not agree to make IPD available (4 trials) and trials were not conducted within a specific time (5 trials); for the remaining 6 trials, no specific reason was provided. Of 619 RCTs published between 2014 - 2016 in 7 high-ranked anaesthesiology journals, only 24 (4%) had a data-sharing statement and none provided data in the manuscript or a link to data in a repository.[23] In a survey targeting the authors of these RCTs, 86 (14%) responded and raw data was obtained from 24 participants. The authors conclude that willingness to share data among anaesthesiology RCTs is very low. From 1 July 2018, clinical trials submitted to ICMJE journals are required to contain a data-sharing statement. The reporting of the statement was investigated in a 2-month period before and after this date.[24] The proportion of articles with a data-sharing statement was 23% (32/137) before and 25% (38/150) after 1st July 2018, while the number of journals publishing data-sharing statements increased from 4/11 to 7/11. Few data-sharing statements complied fully with the ICMJE journal criteria, and the majority did not refer to individual participant data. A total of 300 trials published in 2017-2018 and approximately equally distributed across orthodontics and periodontics were selected, assessed, and analysed with respect to transparency and reporting.[25] Open data-sharing (repository or appendix) was found in 5 % of the trials (11/150 orthodontics and 4/150 periodontics trials). Articles on reproducible research practices and transparency in reproductive endocrinology and infertility (REI) were investigated for original articles with a study type mix from REI journals (2013, 2018) and articles published in high-impact general journals between 2013 – 2018.[26] Raw data was available on request or via online database for 1/98 articles in reproductive endocrinology and infertility RCTs (2013), 0/90 in 2018 and 1/34 in high impact journals. In a random sample of 151 empirical studies in 300 otolaryngology research publications, using a PubMed search for records published between 1 January 2014 and 31 December 2018, only 5 provided a data availability statement and 3 (2.0%) indicated that data was available.[27]

*Metrics of data-sharing statements in clinical trial registries*

Intention to share could be even lower when considering data-sharing plans of trials registered at clinicaltrials.gov. Here the willingness to share data is between 5 and 10%. In one study, 25 551 trial records responded to the Plan to share IPD (72%). Of these, 10.9% of the records indicated "yes" and 25.3% indicated "undecided".[70] Differences were observed by key funder type, with 11% of NIH funders and 0% in the industry answering yes. Importantly, an in-depth review of 154 data-sharing plans suggested a possible misunderstanding of IPD sharing with discrepancies found between data-sharing plans and reports of actual data-sharing. In a survey, the prevalence and quality of IPD-sharing statements among 2,040 clinical trials first posted on ClinicalTrials.gov between 01 January 2018 and 06 June 2018 were investigated.[28] The vast majority of trials included in this study did not indicate an intention to share IPD (n = 1,928; 94.5%). Among the trials that did commit to sharing IPD (n = 112, 5.5%), significant variability existed in the content and structure of the IPD sharing statements with a need for further clarification, enhanced clarification and better outreach. Data from 287.626 clinical trials registered in ClinicalTrials.gov on 20 December 2018 were analysed with respect to sharing of IPD.[29] Overall, 10.8% of trials with a first registration date after December 1 2015 answered "Yes" to plans to share de-identified IPD data. The sharing rate ranged from 0% (biliary tract neoplasms) to 72.2% (meningitis, meningococcal infection) when analysed by disease. For the case of HIV, which was analysed separately, the sharing rate was higher on average (24.5%). In a prediction model, studies that deposit basic summary results on ClinicalTrials.gov, large studies and phase 3 interventional studies are the most likely to declare intention to share IPD data.

*Other data sources*

A 2015 survey,[30] focused on PCORnet (The National Patient-Centered Clinical Research Network), found that a possible barrier toward data-sharing intentions related to how data can be used when shared with institutions that have different levels of experience, and to the possibility of some “competition” between institutions on the marketplace of ideas.

*Experimental studies*

Experimental data suggests that estimations of intention to share data could differ depending on the formulation of the request. For instance, a small randomised prospective study conducted in 2001 including 29 corresponding authors of research publications published in the BMJ, explored their preparedness to share the data from their research.[31] The email contact, randomly allocated, was in one of two forms, a general request (asking if the author would "in general" be prepared to release data for re-analysis) and a specific request (a direct request for the data for re-analysis). Researchers receiving specific requests for data were less likely and slower to respond than researchers receiving general requests. Similarly, in 2019, a randomized controlled trial in conjunction with a Web-based survey[32] included study authors to explore whether and how far a data-sharing agreement affected primary study authors' willingness to share IPD. The response rate was relatively low (21 %) in this study since more than 1,200 individuals were initially contacted and 247 responded. Among the responders, study authors who received a data-sharing agreement were more willing to share their data set, with an estimated effect size of 0.65 (95% CI [0.39, 0.90]).

Authors of published reports on prevention or treatment trials in stroke were asked to provide data for a systematic review and randomised to receive either a short email with a protocol of the systematic review attached ('Short') or a longer email that contained detailed information, without the protocol attached ('Long').[33] 88 trials with 76 primary authors were identified in the systematic review, and of these, 36 authors were randomised to Short (trials=45) and 40 to Long (trials=43). Responses were received for 69 trials. There was no evidence of a difference in response rate between trial arms (Short vs Long, OR 1.10, 95% CI 0.36 to 3.33).

***Trial participants***

*Qualitative studies*

Perceptions of trial participants toward data-sharing and their intention to share were explored qualitatively. A systematic review with a thematic analysis[34] of 9 qualitative studies from Africa, Asia, and North America identified four key themes emerging among patients: the benefits of data sharing (including benefit to participants or immediate community, benefits to the public and benefits to science or research), fears and harm (including fear of exploitation, stigmatization or repercussions, alongside concerns about confidentiality and misuse of data), data-sharing processes (mostly consent to the process), and the relationship between participants and research (e.g. trust in different types of research or organizations, relationships with the original research team). Some qualitative reports provide data on heterogenous samples including patients and various stakeholders from low- and middle-income countries. In-depth interviews and focus group discussions involving 48 participants in Vietnam suggested that trial participants could be more willing to be involved in data-sharing than trialists.[35] A similar study on a range of relevant stakeholders in Thailand[36] found that data-sharing was seen as something positive (e.g. a means to contribute to scientific progress, better use of resources, greater accountability, and more output) but it underlined considerable reservations, including potential harm to research participants, their communities, and the researchers themselves.

In a qualitative study with 16 in-depth interviews, cancer patients currently participating in a clinical trial indicated a general willingness to allow re-use of their clinical trial data and/or samples by the original research team, and supported a generally open approach to sharing data and/or samples with other research teams, but some would like to be informed in this case.[37] Despite divergent opinions about how patients prefer to be involved, ranging from passive contributors to those explicitly wanting more control, participants expressed positive opinions toward technical solutions that allow their preferences to be taken into account.

*Surveys*

Two surveys performed in the US and one in Italy assessed the intention-to-share rates among trial participants (see **Figure 4**). In one survey[38] with a moderate response rate (47%), 463/799 (58%) patients favored or strongly favored data-sharing, while only 9% were against or strongly against it. Most participants (84%) believed that disclosing the data-sharing plan within the informed consent process was important or very important. A higher percentage of ethnic minority participants was against data-sharing (white, 6%, vs. "other", 13).

In a second survey[39] with a high response rate (79%), 93% were very or somewhat likely to allow their own data to be shared with university scientists and less than 8% of respondents felt that the potential negative consequences of data-sharing outweighed the benefits. Predictors of this outcome were a low level of trust in others, concern about the risk of re-identification or about information theft, and having a college degree. 93% and 82 % respectively were very or somewhat likely to allow their data to be shared with academic scientists

and scientists in for-profit companies. The purpose for which the data would be used did not influence willingness to share data except for use in litigation. However, patients were concerned that data-sharing might make others less willing to enroll in clinical trials, that data would be used for marketing purposes, or that data could be stolen. Less concern was expressed about discrimination and exploitation of data for profit.

In a survey of Italian patient and citizen groups, 280/2003 contacts provided questionnaires eligible for analysis.[40] 144/280 (51%) had some knowledge about the IPD sharing debate and 60/280 (42%) had an official position. Of those who had an official position 35/60 (58%) were in favour and 19/60 (32%) in favour with restrictions. 39% approved broad access by researchers and other professionals to identified information.

*Other data sources*

While consent seems to be a crucial issue for trial participants, an analysis of 98 Informed Consent Forms (ICFs) found that only 6 (4%) indicated a commitment to share de-identified IPD with third party researchers.[41] Commitments to share were more common in publicly funded trials than in industry-funded trials (7% vs 3%).

***Publishers/funders***

*Publishers*

*Metrics of data sharing statements and policies*

Several studies were found about the intentions (and data-sharing policies) of publishers. Many publishers have developed data-sharing policies (20-75%), however, less than 10% are mandatory (see **Figure 4**). In a 2009 survey[42] of editors of different member journals of the World Association of Medical Editors (WAME) (response rate 22%), 2% and 19% of journals respectively required provision of participant level data and specification by authors of their data-sharing plan. A similar survey of 10 high-impact surgical journals in 2009 and 2012 found only one journal that had a mandatory data-sharing policy.[43] Data-sharing statements were found only in 2/246 (1%) RCTs published in these 10 journals. Another study of a random sample of 60 journals[44] found that 21 (35 %) provided instructions for patient-level data, but only 4 (7 %) required sharing of IPD (all were oncology journals). A review of 88 websites of dental journals[45] suggested that 17 accepted raw data as complementary material. A 6-year cross-sectional investigation of the rates and methods of data-sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018 was performed.[46] 8/14 (57.1%) journals had data-sharing policies for published RCTs. Of the 394 RCTs included none shared their data publicly. 40/60 clinical psychology journals had a specific policy for data-sharing (2017).[47] Only one journal made data-sharing mandatory, while 37 recommended it. The findings suggest great heterogeneity in journal policies and little enforcement. Online instructions for authors from 38 high-impact addiction journals were reviewed for 6 publication procedures, including data-sharing (2018). 28/38 (74%) of the addiction journals had a data-sharing policy, none was mandatory.[48] It was concluded that many addiction journals have adopted publication policies, but more stringent requirements have not been widely adopted. Instructions for authors in 43 high-impact nutrition and dietetics journals were reviewed with respect to procedures to increase research transparency (2017).[49] 25/33 (75%) journals publishing original research and 4/10 review journals had a data-sharing policy

Among 109 peer-reviewed and original research-oriented dental journals that were indexed in the MEDLINE and/or SCIE database in 2018, a data-sharing policy was present in 32/109 (29.4%) and 2 of these had a mandatory policy.[50] This study concluded that at present data-sharing policies are not widely endorsed by dental journals. In a cross-sectional survey 14 ICMJE-member journals and 489 ICMJE-affiliated journals that published a RCT in 2018 were evaluated with respect to data-sharing recommendations.[51] 8/14 (57%) of member journals and 145/489 (30%) of affiliated journals had an explicit data-sharing policy on their website. In RCTs published in member journals with a data-sharing policy, there were data-sharing statements in 98/100 (98 %) with expressed intention to share individual patient data in 77/100 (77%). In RCTs published in affiliated journals with an explicit data-sharing policy, data-sharing statements were rare 25/100 (25%), and expressed intentions to share individual participant data were found in 22/100 (22%).

Changes in policies from 2013 to 2016 regarding public availability of published research data were investigated in 115 paediatric journals.[52] In 2012 77 /115 (67%) and in 2016 56/115 (49%) accepted storage in thematic or institutional repositories. Publication of data on a website was accepted by 27/115 (23%) and 15/115 (13%). Most paediatric journals recommend that authors deposit their data in a repository but they do not provide clear instructions for doing so.

*Funders and clinical trial units*

*Metrics of data sharing policies by funders*

Several studies investigated mandatory data-sharing policies of funders. 30-80% of the non-commercial funders provided data-sharing policies, the highest rates were observed in the US. Only around 10-20% of these policies were mandatory (see **Figure 4**). In one study 50% of the top non-commercial funders had a data-sharing policy but it was found that in only 2/20 cases data-sharing was required. Six funders offered technical or financial resources to support IPD sharing.[53] Trial transparency policies were investigated for 9/10 top non-commercial funders in the US (May to November 2018).[54] 7/9 (78%) funders had a policy for individual patient data-sharing, for 1 it was mandatory. 6 offered data-sharing and 5 monitored compliance. Of 96 responders out of 190 non-commercial funders contacted in France, 31 were identified as funding clinical trials (2019).[55] 9/31 (29%) had implemented a data-sharing policy. Among these 9 funders, only one had a mandatory sharing policy and 8 a policy supporting but not enforcing data-sharing. Funders with a data-sharing policy were small funders in terms of total financial volume.

Three studies investigated mandatory data sharing policies among commercial sponsors (see **Figure 4**). In a 2016 survey, 22/23 (96%) companies among the top 25 companies by revenue had a policy to share IPD[59]. In a second sample of 42 unselected companies, 30 (71 %) had one. These policies generally did not cover unlicensed products or trials for an off-label use of a licensed product. 52 % of top companies, and 38 in the sample including all companies considered that requests for IPD for additional trials were not explicitly covered by their policy.[56] A second survey[57] studied data availability for 56 publications reporting on 61 industry-sponsored clinical trials of medications. Of these 61 studies, 32 (52%) had a public data-sharing policy/process.

78 non-commercial funders and a sample of 100 leading commercial funders in terms of drug sales having funded at least one RCT in the years 2016 to 2018 were surveyed (15 February 2019 – 10 September 2019).[58] 30/78 (38%) non-commercial funders had a data-sharing policy with 18/30 (60%) making data-sharing mandatory and 12/30 (40%) encouraging data-sharing. 41/100 (41%) of the commercial funders had a data-sharing policy. Among funders with a data-sharing policy, a survey of two random samples of 100 RCTs registered on Clinicaltrial.gov found that data-sharing statements were present for 77/100 (77%) and 81/100 (81%) of RCTs funded by non-commercial and commercial funders respectively. Intention to share data was expressed in 12/100 (12%) and 59/100 (59%) of RCTs funded by non-commercial and commercial funders. The survey indicated suboptimal performance by funders in setting up data-sharing policies.

*Metrics of data-sharing policies by CTUs*

Among 23 UK Clinical Research Collaboration (UKCRC) registered Clinical Trial Units (CTUs)[10] (response rate = 51 %), 5 (22 %) had an established data-sharing policy and 8 (35%) specifically required consent to use patient data beyond the scope of the original trial (see table). Concerns were raised about patient identification, misuse of data, and financial burden. No CTUs supported the use of an open access model for data-sharing.

*Other data sources*

A 2005 survey[59] of 107/122 accredited medical schools in the US (response rate = 88%) explored data-sharing in the context of contractual provisions that could restrict investigators' control over data in the context of industry-funded trials. There was poor consensus among senior administrators in the offices of sponsored research at these institutions on the question of prohibiting investigators from sharing data with third parties after the trial is over (41 % allowed it, 34 % disallowed it, and 24 % were not sure whether they should allow it).

In a survey targeting European heads of imaging departments and speakers at the clinical trials in radiology sessions (July – September 2018), the response rate was 132/460 (29%).[60] Responses were received from institutions in 29 countries, reporting 429 clinical trials. For future trials, 98% of respondents (93/95) said they would be interested in sharing data, although only 34% had already shared data (23/68). The main barriers to data-sharing were data protection, ethical issues, and lack of a data-sharing platform.

**Results for individual sources of evidence: actual data-sharing**

***Re-users***

*Studies related to journal articles*

*Metrics of actual data-sharing*

Several studies have been performed investigating data-sharing rates for studies that have been published in journals, the majority with data-sharing policies and high impact (**Figure 5**). Even with strict data-sharing policies, the data-sharing rates are low or at most moderate, and vary between 10 and 46%, except for one study with a very high data-sharing rate due to a partly preselected sample of authors willing to share their data[18] . In the 6-year cross-sectional investigation of the rates and methods of data-sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018, none of the 394 clinical trials included shared their data publicly.[46] Of 86 responders in a survey targeting the corresponding authors of 619 RCTs published between 2014 - 2016 in 7 high-ranking anaesthesiology journals, raw data was obtained only for 24 studies.[23] 62 declined to share raw data. In a study targeting PLOS Medicine and PLOS Clinical Trials publications conducted in 2009, 1/10 (10%) of the data sets was made available after request[28]. In articles in Chinese and international journals from 2016, sharing practices were indicated for 29/247 (11%) of the articles.[19] Among the top 10 general and internal medical journals investigated in 2016, IPD was provided after request for 9/61 (15%) of pharmaceutical-sponsored studies [56]. For BMJ research articles published between 2009 and 2015, data sets were made available in 7/157 (4%) of the articles.[30] For the sub-sample of clinical trials the rate was higher (5/21 (24%)). Of 317 clinical trials published in 6 general medical journals between 2011 and 2012, 115 (36%) granted access to data[35] . The data availability for RCTs published in BMJ and PLOS Medicine between 2013 and 2016 was 17/37 (46%)[42].

*Experimental studies*

In a parallel group RCT, an intervention group (offer of an Open Data Badge for data-sharing) was compared to a control group (no badge for data-sharing).[61] The primary outcome was the data-sharing rate. Of 160 research articles published in BMJ Open, 80 were randomised to the intervention and control groups, of which 57 could be analysed in the intervention group and 54 in the control group. In the intervention group data was available on a third-party repository for 2/57 (3.5%) and upon request for 32/57 (56.1%) respectively in the control group: 3/54 (5.6%) and 30/54 (56%). Data-sharing rates were low in both groups and did not differ between groups.

*Data sharing for IPD meta-analyses*

*Metrics of data-sharing for IPD meta-analyses*

Some examples demonstrate that data availability for IPD meta-analyses is still limited despite the various data-sharing initiatives/platforms (**Figure 5**). Availability can be increased under specific circumstances, such as the creation of a disease-specific repository for a scientific community, as demonstrated for a repository of IPD from multiple low back pain RCTs with IPD from 20/42 (48%) RCTs included[57], and a study on anti-epileptic drugs conducted by a Cochrane group with IPD for 15/39 (38%) studies included[40]. In another study on different databases, 35 individual participant data meta-analyses with more than 10 eligible RCTs were identified (May 1, 2015 to February 13, 2017).[61] Of 774 eligible RCTs identified in these meta-analyses, 517 (66.8 %) contributed data. The country where RCTs are conducted (the UK versus the United States (US)), the impact factor of the journal (high versus low) and a recent RCT publication year were associated with higher sharing rates. In three other studies, the availability of datasets for IPD meta-analysis was limited (0-17%). In one study performed in 2014, devoted to one commercial sponsor with one specific medicinal product, IPD from 24 trials was requested without success[47] Of 15 requests (13 direct to authors, 2 to a repository) in 2014/2016, IPD was received for 2/15 (13%) of the studies[51]. Of 217 RCTs published since 2000 in orthopaedic surgery, agreement to send IPD was obtained for 37/217 (17%)[35].

*Experimental studies*

The low data availability for IPD-meta-analyses is underlined by two experimental studies. One experimental study covered the issue of actual data-sharing. In this small randomized prospective study,[31] where 29 corresponding authors of original research articles in a medical journal were contacted via two different modes (general versus specific request), only one author actually sent the data immediately in response to a specific request and one author, without caveats, reported willingness to send the data in response to a general request.

A randomized controlled trial investigated the effect of financial incentives on IPD sharing.[62] All study participants (129 in all) were asked to provide the IPD from their RCT. Those allocated to the intervention group received financial incentives, those from the control group did not. The primary outcome was the proportion of authors who provided IPD. None of the authors shared their IPD, whichever the group.

*Other data sources*

Two studies investigated the completeness of data availability in IPD meta-analyses. Out of 30 IPD meta-analyses included in a survey,[63] 16 did not have all the IPD data requested. The access rate for retrieving IPD for use in IPD-meta-analyses was investigated in a systematic review.[64] Only 188 (25%) of 760 IPD meta-analyses retrieved 100% of the eligible IPDs for analysis and there was poor evidence that IPD retrieval rates improved over time.

*Access to repositories/platforms*

Only a few studies describe access to repositories/platforms from the viewpoint of the user (**Figure 5**). Experiences with two major platforms (CSDR, PDS) were reported.[65] In these very early-phase projects, no data access was possible with CSDR, and faster data acquisition was achieved via the Project Data Sphere. High sharing rates were reported for academic repositories (MRC CTU, BioLINCC), Of 103 requests to MRC CTUs, access was granted in 80/103 (78%) cases[22] . In a survey of investigators 536/536 (100%) received access to BioLINCC over a time period between 2007 and 2014[31].

***Repositories/platforms***

*Commercial sponsors*

*Metrics of actual re-use*

Different initiatives and platforms were initially implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now open to academic trials but this has not been used very often so far). This covers the YODA project, CSDR, Vivli and SOAR (which is now part of Vivli). For the different platforms and repositories, metrics describing the actual use of data are available (**Figure 5**).

6 studies have accessed data-sharing rates for CSDR. From 2014 to the end of January 2019, there was a total of 473 research proposals submitted to CSDR.[66] Of these, 364 met initial administrative and data availability checks, and the independent review panel approved 291. 222/473 (46.9%) of the requests gained access to the data (in progress and completed). Of the 90 research teams that had completed their analyses by January 2018, 41 reported at least one resulting publication to CSDR. Less than half of the studies ever listed on CSDR have been requested. Between 2014 and 2017 CSDR received a total of 172 research proposals, of which 105 (61%) were approved[26]. In another study focusing on availability and use of shared data from cardiometabolic clinical trials in CSDR covering the time period between 2013 and 2017, 198 (62%) were approved with or without conditions[18]. In year one of the use of CSDR (2013-2014), 36 research proposals were approved with conditions, of these 23 (64%) progressed to a signed data-sharing agreement[24]. From 2014 to 2017, Boehringer-Ingelheim listed 350 trials for potential data-sharing at CSDR.[67] 55 research proposals were submitted, of which 37 (67.3%) were approved. All approved research proposals submitted to Boehringer-Ingelheim except one addressed new scientific questions or were structured to generate new hypotheses for further confirmatory research, rather than replicating analyses by the sponsor to confirm previous research. Between 2013 and 2015 177 research proposals were submitted to CSDR, and access was granted for 144 (81%) of these proposals[23].

In the first year following the launch in October 2014, YODA received 29 requests all of which were approved (100%) [49]. In 2017 the YODA project reported 73 proposals of which 65 were approved (Ross, 2017). A more recent publication reported the metrics for data-sharing of Johnson & Johnson clinical trials in the YODA project up to August 27, 2018.[68] 100 data requests were received from 89 principal investigators (PI) for a median of 3 trials per request. 90/100 requests (90 %) were approved and a data use agreement was signed in 82/100 (82%).

The use of the open access platforms CSDR, ODA and SOAR together between 2013 and 2015 was investigated in one study. Of the 234 proposals submitted, 154 (66%) were approved[41]

The data available shows that the use of these platforms has increased steadily since their initiation and that 50% and more of the data requests lead to actual data-sharing. The reasons for not sharing are numerous but data access is rarely denied by the platforms. Our assessment of CSDR, YODA, NIDDK and Vivli websites is presented in **Table 1**.

**Table 1**: Metrics of CSDR, YODA, and Vivli websites
NIDDK also provided metrics concerning the number of requests (530) but no other information
*publication anticipated

| Platform | Metrics date | Available studies | Number of requests | Number of requests with data shared | Number of requests with data leading to publication | Number of publications |
|---|---|---|---|---|---|---|
| CSDR | 30/11/2020 | 3008 | 621 | 318 | 59* | 79 |
| YODA | 15/11/2019 | 334 | 196 | 173 | 29 | 35 |
| Vivli | 02/11/2020 | 5203 | 215 | 123 | 8 | 9 |

*Metrics of trial coverage for data-sharing*

Ethics approval in applications for open-access clinical trial data from CSDR was investigated in a survey.[69] Projects with and without ethics approval were applied to at roughly similar rates (62/111 and 43/61).
The proportion of trials where the pharmaceutical and medical device industry provided IPD for secondary analyses and thus the completeness of trial data is still limited.[57] Only 15% of 61 industry-sponsored clinical trials were available 2 years after publication. For companies listing at least 100 studies on CSDR, a search was performed in ClinicalTrials. gov (1/2016, studies terminated/ completed at least 18 months before search date).[70] Among 966 RCTs registered in ClinicalTrials.gov, only 512 (53%) were available on CSDR and only 385 (40%) of the RCTs were registered and listed on CSDR with all datasets and documents available. This was the case despite the time lapse of 18 months since the completion of the drug trials by the company sponsor. Differences across sponsors were observed. Pharmaceutical repositories may cover only part of the trials with commercial sponsors needed for meta-analyses. In a study investigating data availability for industry-sponsored cardiovascular RCTs with more than 5000 patients, performed by a top-20 pharmaceutical company and registered at ClinicalTrials.gov (up to Jan. 2015), only 25% of the identified trial data was confirmed to be available.[22] In 50% of cases availability could not be definitely confirmed.
As part of the Good Pharma Scorecard project, data-sharing practices were assessed for large pharmaceutical companies with novel drugs approved by the FDA in 2015, using data from ClinicalTrials.gov, Drugs@FDA, corporate websites, data-sharing platforms and registries (e.g. YODA, CSDR).[71] 628 trials were analysed. 25% of the large pharmaceutical companies made IPD accessible to external investigators for new drug approvals, this proportion improved to 33% after applying a ranking tool.

*Non-commercial sponsors*

Disease-specific academic clinical trial networks have a long history of IPD sharing, especially US-related NIH institutions. This is clearly demonstrated by the available literature; however, the metrics of data-sharing are

not always as transparent as with the industry platforms, and data cannot be structured and documented easily in a table.

In a survey on the use of the National Heart, Lung, and Blood institute Data Repository, access to 100 studies initiated between 1972 and 2010 was investigated.[72] A total of 88 trial datasets were requested at least once, and the median time from repository availability and the first request was 235 days.

Since its inception in 2006 and through to October 2012, nearly 1700 downloads from 27 clinical trials have been accessed from the Data Share website belonging to the National Drug Abuse Treatment Clinical Trial Network (CTN) in the US, with use increasing over the years.[73] Individuals from 31 countries have downloaded data so far.

In a case study approach, the data-sharing platform Data Share of the National Institute of Drug Abuse (NIDA) was investigated in detail.[74] As of March 2017, the Data Share platform had included 51 studies from two trial networks (36 studies from CTN and 15 studies from NID Division of Therapeutics and Medical Consequences). From 2006 to March 2017, there have been 5663 downloads from the Data Share website. Of these, 4111 downloads have been from the US.

The Project Data Sphere (PDS) is an open-source data-sharing model that was launched in 2014 as an independent, non-profit initiative of the CEO roundtable on cancer.[75] PDS contains data from 72 oncology trials, donated by academics, governments, and industry sponsors. More than 1400 researchers have accessed the PDS database more than 6500 times. As an example, a challenge to create a better prognostic model for advanced prostate cancer was issued in 2014, with 549 registrants from 58 teams and 21 countries.

The Immune Tolerance Network (ITN) is a National Institute of Allergy and Infectious Diseases /National Institutes of Health-sponsored academic clinical trial network.[76] The trial sharing portal, which was released for public access in 2013, provides complete open access to clinical trial data and laboratory studies from ITN trials at the time of the primary study publication. Currently, data from 20 clinical trials is available and data for an additional 17 will be released to the public at the timepoint of first publication. So far, more than 1000 downloads have been registered.

In the MRC Clinical Trials Transparency Review Final Report (November 2017), the MRC United Kingdom (UK) reported that 24/107 (22%) trials that started during the review period had created a database for sharing. Seven of these datasets (7/24, 29%) had already been shared with other researchers.[77]

Of 215 requests submitted for PLCO data, 199 (93%) were approved, and for NLST 214 (89%) out of 240 requests.[78]

*Other stakeholders*

In a case study about experiences with data-sharing among data monitoring committees, access to five concurrent trials assessing the level of arterial oxygen, which should be targeted in the care of very premature neonates, was investigated.[79] The target of taking all relevant evidence into account when monitoring clinical trials could be only partially reached.

One case-study directly addressed the issue of costs. Data from two UK publicly funded trials was used to assess the resource implications of preparing IPD from a clinical trial to share with external researchers.[80] One trial, published in 2007, required 50 hours of staff time with a total estimated cost of £3185, and the other published in 2012 required 39.5 hours with £2540.

**Results of individual sources of evidence: re-use**

***Any type of re-use***

The majority of research projects using shared clinical trial data are dealing with new research. This covers studies on risk factors and biomarkers, methodological studies, studies on optimizing treatment and patient stratification and subgroup analyses. IPD meta-analyses were a less frequent reason for data-sharing requests to repositories and only a few have been reported. Re-analyses are only exceptionally applied.

Early experiences with CSDR, involving GlaxoSmithKline trials[81] found low rates of IPD meta-analyses and re-analyses, the vast majority being secondary analyses (studies on risk factors or biomarkers, methodological studies, predictive toxicology or risk models, studies of optimizing treatments, subgroup analyses etc.). Similar results were found in an update of the analysis.[82]

In the YODA project, which had received 73 proposals for data-sharing as of June 2017 and had approved 65 proposals,[83] the most common study purposes were to address secondary research questions (n=39), to combine data as part of larger meta-analyses (n=35) and/or to validate previously published studies (n=17).

Among the 172 requests to the National Heart, Lung and Blood Institute (NHLBI) data repository with online project descriptions and coded purpose, 72% of requests were initiated to address a new question or hypothesis, 7% to perform a meta-analysis or combined study analysis, 2% to test statistical methods, 9% to investigate methods relevant to clinical trials, and 9% for other reasons.[72] In only two requests, the available description suggested a re-analysis.

From 2014 to the end of January 2019, 222/473 (46.9%) of the requests to CSDR gained access to the data (in progress and completed).[66] 90/222 (40.5 %) of the research teams had completed their analyses by January 2018. 41 published at least one paper, and another 28 that were expected to publish shortly.

In the SPRINT challenge, individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and to identify novel scientific or clinical findings.[84] Among 200 qualifying teams, 143 entries were received.

***Further additional analyses***

There were few indications concerning the exact type of secondary analysis that was performed. Approved proposals per subject matter are available for the Cancer Data Access system (CDAS), covering two large cancer screening trials (Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial and National Lung Screening Trial (NLST).[78] Of the 199 approved requests to PLCO between November 2012 and October 2016, 84 (42%) were devoted to cancer etiology, 66 (33%) to trial-related screening, 29 (15%) to other areas, 14 (7%) to risk prediction and 6 (3%) to image analysis. Of the 214 approved requests to NLST, 95 (44%) were devoted to image analysis, 90 (42%) to trial-related screening, 14 (7%) to other subjects, 10 (5%) to cancer etiology and 5 (2%) to risk prediction.

***IPD meta-analyses***

In one study, IPD meta-analyses proved to amount to a small proportion of data re-use. Among the 174 research proposals approved up to 31 August 2017 by CSDR, 12 proposals were IPD meta-analyses, including network meta-analyses.[85] All were retrospective IPD meta-analyses (i.e. none was a prospective IPD meta-analysis).

***Re-analyses***

A 2014 survey of published re-analyses[86] found that a small number of reanalyses of RCTs have been published (only 37 re-analyses of 36 initial RCTs) and only a few were conducted by entirely independent authors. 35% of these reanalyses led to changes in findings that implied conclusions different from those of the original article for the types and numbers of patients who should be treated.

In the survey of 37 RCTs in the BMJ and PLOS Medicine[87] published between 2013 and 2016, 14 out of 17 (82%, 95% IC: 59% to 94%) available studies were fully reproduced on all their primary outcomes. Of the remaining RCTs, errors were identified in two, but reached similar conclusions, and one paper did not provide enough information in the Methods section to reproduce the analyses.

**Results for individual sources of evidence: output from data sharing**

Publications can be considered as the main research output of data-sharing. Publication activity in the re-use of clinical trial data was considered in several studies. Detailed data are available for academic clinical trial networks and disease-specific repositories in the US, some of them already practising data-sharing for a period longer than 10 years. Here, fair to moderate publication output has been observed depending on the individual repository. So far this is not the case for the repositories storing clinical trial data from commercial sponsors,

taking into consideration that these repositories were established around five years ago and that there is usually a considerable time lag between request, approval, analysis and publication. Current statistics indicate improvement in publication output with time.

***Non-commercial sponsors***

In a cross-sectional web-based survey about access to clinical research data from BioLINCC, covering the period from 2007 to 2014, 98 out of 195 responders (50%) reported that their projects had been completed, among which 66 (67%) had been published.[88] Of the 97 respondents who had not yet completed their proposed projects, 81 (84%) explained that they planned to complete their project; 63 (65%) indicated that their project was in the analysis/manuscript draft phase.

In a survey targeting European heads of imaging departments and speakers at the Clinical Trials in Radiology sessions (July – September 2018), 23/68 reported that they had already shared data.[60] At least 44 original studies were published based on the data shared by the 23 institutions involved.

In five studies (**Table 2**) the number of publications was reported, usually referring to the number of trials included in the repository/platform.

**Table 2**: Studies reporting published outputs for non-commercial sponsors

| Reference | Repository/ platform | No. of trials included in repository/platform | No. of published articles | Assessment |
|---|---|---|---|---|
| Shmueli-Blumberg, 2013 | CTN Data Share | 27 trials (1700 downloads) | 13 | 2012 |
| Zhu, 2017 | CDAS | 2 trials (PLCO, NLST) (455 requests) | 25% for PLCO projects, 19% for NLST projects | 2016 |
| Coady, 2017 | BioLINCC | 100 trials (88 requested at least once) | 35% of clinical trials at least 1 publication 5 years after availability in the repository | 5/2016 |
| Huser, 2018 | NIDA Data Store | 51 trials | 14 | 3/2017 |
| Pisani, 2017 | WWARN | 186 trials | 18 | 2016 |

***Commercial sponsors***

Various studies explored metrics of both YODA and CSDR (**Supplementary Material 4**).

Up to 2021, Vivli's website indicates very little published output. We were not able to retrieve published output from NIDDK. **Figure 6** presents publication metrics for CSDR (up to 31 August 2019) and YODA (up to 1st July 2019). Among 88 published papers (62 from CSDR and 26 from YODA), 49 were secondary analyses (42 from CSDR and 7 from YODA), 30 were meta-analyses (13 from CSDR and 17 from YODA), 6 were methodological studies (5 from CSDR and 1 from YODA) and 3 were re-analyses (2 from CSDR and 1 from YODA). The details of these publications[82 83 89] are presented in **Supplementary Material 5**.

**Results of individual sources of evidence: impact of research output**

Evidence on the impact of research output from sharing IPD from clinical trials is still very sparse. So far only two studies, with inconsistent results dealing with this issue and focusing only on citation metrics could be identified.

*Metrics on citations*

One study, already published in 2007, suggested that sharing detailed research data was associated with an increased citation rate.[90] Of 85 cancer microarray clinical trials published between January 1999 and April 2003 41 made their microarray data publicly available on the internet. For 2004 – 2005, the trials with publicly available data received 85% of the aggregate citations. Publicly available data was significantly associated with a 69% increase in citations, independently from journal impact factor, date of publication and the author's country of origin.

Citation metrics for 224 publications based on repository data for clinical trials in the NHLBI Data Repository were compared with publications that used repository observational study data, as well as a 10%-random sample of all NHLBI-supported articles published in the same period (January 2000 – May 2015).[72] Half of the publications based on clinical trial data had cumulative citations that ranked in the top 34% normalized for subject category and year of publication, compared to 28.3% for publications based on observational studies and 29% for random samples. The differences were, however, not statistically significant.

*Other data sources*

In the SPRINT challenge, individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and to identify novel scientific or clinical findings.[84] Among 200 qualifying teams, 143 entries were received. Entries were judged by a panel of experts on the basis of the utility of the findings to clinical medicine, the originality and novelty of the findings, and the quality and clarity of the methods used. All submissions were also open for crowd voting among the 16,000 individuals following the SPRINT Challenge. Cash prizes were awarded, and winners were invited to present their results. 143 entries to the SPRINT data challenge were received.

## DISCUSSION

### Summary of evidence

There are major differences with respect to the intention to share IPD from clinical trials across the different stakeholder groups. The studies available so far show that clinical trialists and to some extent study participants, as the two main actors of clinical trials, usually have great willingness to share data (60-80%). This is much less pronounced when it comes to data-sharing statements published in journal articles. Depending on the journals considered, the rates vary from less than 5% to around 25%. The situation is even worse when data-sharing plans documented in registries (e.g. ClinicalTrials.gov) are analysed. Here the willingness to share data is between 5 and 10%.

As a consequence, considerable discrepancy between the positive attitude towards data-sharing in general and the intention to do so in an actual study needs to be ascertained. Publishers, enabling the publication of research output from clinical trials and funders/sponsors financing clinical trials, could be major drivers to change the situation. Meanwhile many publishers have developed data-sharing policies (20-75%), but less than 10% are mandatory and have thus not been enforced. There are differences between journals, with some of the high-impact journals being more involved in the data sharing movement than the others *(e.g PLOS Medicine, the BMJ, Annals of Internal Medicine).* For funders, the situation is similar, but differs between commercial and non-commercial funders. 30-80% of the non-commercial funders provide data-sharing policies, with the US and NIH at the front. Only around 10 to 20% of these policies are mandatory. Data-sharing policies have been developed more often in the group of commercial funders (40-95%) but information on the proportion of mandatory policies is lacking. In short, the pressure by publishers and funders to share data is still limited and the situation is only slowly improving. The situation is better for the pharmaceutical industry, which has not only promoted data-sharing policies in their organisations to a large degree but has also implemented platforms and repositories, providing practical support for the process of data-sharing (e.g. CSDR, Yoda, Vivli).

Several studies have been performed investigating data-sharing rates for clinical studies that have been published in journals. The focus has been on high-impact journals with strict data-sharing policies (e.g. PLOS Medicine, BMJ, Ann Intern Med), demonstrating data-sharing rates between 10% and 46%, except for one

study with a very high data-sharing rate due to a partly preselected sample of authors willing to share their data. Data availability for IPD meta-analyses is usually limited (0-20%), available only under specific circumstances (Cochrane group, disease-specific repository) and the availability can be increased to 50% and more. A few individual studies describe access to repositories/platforms from the viewpoint of the user, which does not enable identification of a general pattern. Different initiatives and platforms have been implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now open to academic trials, but this has not been used very often so far). This covers the YODA project, CSDR, Vivli and SOAR (which is now part of Vivli). The data available shows that the use of these platforms has increased steadily since their initiation and that 50% and more of the data requests lead to actual data-sharing. The reasons for not sharing are numerous but data access is rarely denied by the platforms.

The majority of research projects using shared clinical trial data deal with new research. This covers studies on risk factors and biomarkers, methodological studies, studies on optimizing treatment and patient stratification and subgroup analyses. This is important because new research may be easier to publish in peer-reviewed journals, which is a major driver of academic careers.

So far only some IPD meta-analyses have been planned as part of data-sharing initiatives, and only a few have been reported. There are many hurdles for IPD meta-analyses, including the findability, the accessibility and the re-usability of datasets (F, A and R in FAIR). ECRIN has developed a metadata dictionary (MDR), able to identify clinical studies and data objects related to it (e.g. protocol, DMP, CRF).[91] This tool allow for identifying studies for which datasets are available and the conditions for access (ECRIN, MDR). Even if IPD datasets are accessible for meta-analyses, the studies are usually distributed across various repositories. This has been demonstrated in several studies in our scoping review. One central repository could simplify the situation, but instead, the number of repositories is steadily increasing.[2] The situation could be considerably improved with more standardisation and harmonisation of data and procedures and a federating approach between repositories.

Re-analysis of clinical trial data could help the scientific community to enhance the validity of reported trial results. An illustration is the “restoring study 329” initiative, investigating efficacy and harm of paroxetine and imipramine in the treatment of major depression in adolescence. The re-analysis reached different conclusions with important implications for both clinical practice and research.[3] RIAT (Restoring invisible & abandoned trials support center) was initiated as an international effort to tackle bias in the way research is reported with the goal of providing more accurate information to patients and other healthcare decision makers.[92]

One of the problems that is tackled by RIAT is misreporting (inaccurately or incompletely reported trials). In our scoping review we found that re-analyses are only exceptionally applied. In one review, the majority of studies was reproduced on all primary outcomes, in another around one third of studies led to changes in findings different from the original articles. It seems that re-analysis is only attractive in a minority of cases deserving major public interest. Nevertheless, for these cases, repositories holding and sharing IPD could be very useful and speed up the process of data-sharing. It could be of interest to establish a link between RIAT and data-sharing platforms and initiatives.

Publications can be considered as the main output from data-sharing. Usually, there is a considerable time lag between requesting data for re-use, receiving shared data, performing secondary analysis, writing a manuscript and publishing the secondary analysis. This has to be taken into consideration when the publication output of data-sharing initiatives and platforms is analysed. Repositories and platforms mainly devoted to commercial trials have now existed for around 5 years, so only a limited publication output can be expected. Fortunately, these repositories provide detailed metrics for data-sharing requests, including number and type of publications originating from data-sharing. As expected, the number of publications related to data-sharing for commercial studies is still limited, but current statistics indicate improvement over time. The situation with non-commercial sponsors is different. Academic clinical trial networks and disease-repositories have been successfully implemented (mainly in the US) and have already practised data-sharing for quite a long time, some for more than 10 years. Here data-sharing is part of the research culture and the exchange of data is based on elements such as trust, technical support and common benefit. Outstanding examples are BioLINCC,[88] NIDA[73] and World Wide Antimalarial resistance Network (WWARN).[93] This is reflected in the data-sharing rates for IPD meta-analyses, which are rather low if data requests target authors directly, compared to data-sharing requests within communities (e.g. Cochrane groups) or related to specific repositories. Outside clinical trial networks and disease-specific repositories, data-sharing of IPD is still very limited. Possible reasons could

include the lack of widely accepted repositories for non-commercial clinical trials and insufficient incentives and benefits related to data-sharing. Some investigators may be reluctant to share their data, other may simply not know how to proceed.

We describe secondary analyses as a very popular type of reuse. These analyses are however exploratory and carry a risk of alpha inflation (due to multiple comparisons). Not all results of these analyses have been published. Alpha inflation and selective reporting can be fertile ground for non-reproducible science and this phenomenon surely deserves attention. Improvements could be achieved with a prospective registration of any protocol for secondary data use similar to the trial registries (e.g. ClinicalTrials.gov), a mandatory link between the registration and the original publication or data set and the need to refer to the primary publication or dataset if the re-analysis is published. Existing approaches and tools could then be extended to automatically identify publications related to re-use of data and establish a link to the original work (e.g. see crossmark – crossref[94] , metadata repository (MDR) developed by ECRIN linking clinical studies with related data objects).[91] Another possibility could be to set up a register for secondary analyses.

To be widely accepted, research output from shared data should have an impact on medical research (e.g. generation of new hypotheses) and medical health (e.g. changes in treatment via guidelines). It is well known that the impact of primary studies on medical research and health often has a considerable time-lag and direct effects are not easy to demonstrate. So it is to be expected that evidence from research output from shared data is even more difficult to demonstrate. In this scoping review, taking into consideration the limited time available for data-sharing activities to generate an impact, no major effects were to be expected. As a consequence, the evidence on the impact of data-sharing is still very sparse. This could mean that it is still too early to measure any impact, or that the impact is very limited. So far, only surrogate measures have been considered (citation metrics) with inconclusive results. It is hoped that in the coming years, more studies with more relevant criteria and metrics will be performed. One option could be to closely follow up the SPRINT challenge, where 143 secondary analyses on a single clinical trial were performed, and it would be interesting to see whether one or more of these secondary analyses really had an impact.

**Limitations**

Retrieving and synthesizing information for this study proved to be difficult because we operated in a very siloed landscape where each initiative platform operates with its own metrics. We have tried to be exhaustive by reviewing both the literature and the most important initiatives. However, it was hard to keep the review up-to date as we were studying a moving target in a rapidly changing environment with more and more new initiatives. Some pharmaceutical companies may operate in their own environment and not on larger data-sharing platforms. This makes these activities even more difficult to track. In addition, data-sharing has not had a long history and many of the initiatives and activities were launched in the recent past. Therefore, only a limited research output from data-sharing can be expected so far and indeed, the number of publications is disappointing. It is expected that the number of publications will increase, and indeed we are already seeing this.

**Conclusions**

There is currently a gap in the evidence base evaluating impact of IPD sharing, which causes uncertainties in the implementation and adoption of current data-sharing policies. Data-sharing faces many challenges including, for instance, the scepticism of trialists.[95] There is therefore a need to provide high-level evidence that the value of medical research liable to inform clinical practice increases with greater transparency, and with the opportunity for external researchers to re-analyse, synthesize, or build on previous data. First, a register (such as PROSPERO[96]) for any secondary use of shared data should be created. The inclusion in such a register could be mandatory for any data-sharing agreement/publication, as for the registration of clinical trials. This register would make it possible to build an observatory of data-sharing practices providing direct feedback, without the present silos we have to face. In addition, a register of this sort could help to prevent any selective publication of secondary analyses. Lastly, we suggest that interventional studies should be run to determine the optimal data-sharing policy and/or incentives that add value to clinical research. We do however need to take into consideration that the experimental studies performed so far were not very conclusive, indicating that experimental studies in this area are very demanding.

**REFERENCES**

1. Nosek BA, Alter G, Banks GC, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. *Science (New York, NY)* 2015;348(6242):1422-5. doi: 10.1126/science.aab2374 [published Online First: 2015/06/27]
2. Banzi R, Canham S, Kuchinke W, et al. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials* 2019;20(1):169. doi: 10.1186/s13063-019-3253-3 [published Online First: 2019/03/17]
3. Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. [published Online First: 2015/09/18]
4. Data sharing and the future of science. *Nature Communications* 2018;9(1):2817. doi: 10.1038/s41467-018-05227-z
5. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data: A Proposal from the International Committee of Medical Journal Editors. *PLoS Med* 2016;13(1):e1001950. doi: 10.1371/journal.pmed.1001950 [doi]
PMEDICINE-D-15-03676 [pii] [published Online First: 2016/01/21]
6. Rosenbaum L. Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp. *N Engl J Med* 2017;376(23):2201-03. doi: 10.1056/NEJMp1704482 [doi] [published Online First: 2017/04/27]
7. Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *PLoS Med* 2017;14(6):e1002315. doi: 10.1371/journal.pmed.1002315 [doi]
PMEDICINE-D-17-01550 [pii] [published Online First: 2017/06/06]
8. Vasilevsky NA, Minnier J, Haendel MA, et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* 2017;5:e3208. doi: 10.7717/peerj.3208 [published Online First: 2017/05/04]
9. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7(12):e018647. doi: 10.1136/bmjopen-2017-018647 [published Online First: 2017/12/17]
10. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. *Journal of clinical epidemiology* 2016;70:17-25. doi: 10.1016/j.jclinepi.2015.07.002 [published Online First: 2015/07/15]
11. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 2016;3:160018. doi: 10.1038/sdata.2016.18 [published Online First: 2016/03/16]
12. Peters MD, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare* 2015;13(3):141-6. doi: 10.1097/xeb.0000000000000050 [published Online First: 2015/07/03]
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of internal medicine* 2018;169(7):467-73. doi: 10.7326/m18-0850 [published Online First: 2018/09/05]
14. Clinicaltrials.gov. Learn About Clinical Studies [Available from: https://clinicaltrials.gov/ct2/about-studies/learn, accessed 2021/07/01
15. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of clinical epidemiology* 2016;75:40-6. doi: 10.1016/j.jclinepi.2016.01.021 [published Online First: 2016/03/24]
16. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ (Clinical research ed)* 2012;345:e7570. doi: 10.1136/bmj.e7570 [published Online First: 2012/11/22]
17. Rathi VK, Strait KM, Gross CP, et al. Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials* 2014;15:384. doi: 10.1186/1745-6215-15-384 [published Online First: 2014/10/04]

18. Tannenbaum S, Ross JS, Krumholz HM, et al. Early Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators. *Annals of internal medicine* 2018;169(8):586-88. doi: 10.7326/m18-0723 [published Online First: 2018/07/18]
19. Ji Y, Chen L, Hao X, Wang N & Bao Y. Endorsement of Data Sharing by Authors and High-Impact Medical Journals in China: A Survey of Authors and Assessment of Journal Online Instructions. The 8th International Congress on Peer Review and Biomedical Publication; 2017; Chicago.
20. Griswold M, Laine C, Mulrow C, Schaeffer MB, Stack, C. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. The 6th International Congress on Peer Review and Biomedical Publication; 2009; Vancouver.
21. Rowhani-Farid A, Barnett AG. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ open* 2016;6(10):e011784. doi: 10.1136/bmjopen-2016-011784 [published Online First: 2016/10/16]
22. Murugiah K, Ritchie JD, Desai NR, et al. Availability of Clinical Trial Data From Industry-Sponsored Cardiovascular Trials. *Journal of the American Heart Association* 2016;5(4):e003307. doi: 10.1161/jaha.116.003307 [published Online First: 2016/04/22]
23. Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *Journal of clinical epidemiology* 2019;109:111-16. doi: 10.1016/j.jclinepi.2019.01.012 [published Online First: 2019/02/10]
24. Kaufmann I, Gattrell WT. Assessment of journal compliance with data sharing guidelines from the International Committee of Medical Journal Editors (ICMJE) / in Original Abstracts from the 2019 European Meeting of ISMPP. *Current Medical Research and Opinion* 2019;35(sup2):31-41. doi: 10.1080/03007995.2019.1583496
25. Papageorgiou SN, Antonoglou GN, Martin C, et al. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. *Journal of orthodontics* 2019;46(2):101-09. doi: 10.1177/1465312519842315 [published Online First: 2019/05/09]
26. Kemper JM, Rolnik DL, Mol BWJ, et al. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. *Fertility and sterility* 2020;114(6):1322-29. doi: 10.1016/j.fertnstert.2020.05.020 [published Online First: 2020/08/11]
27. Johnson AL, Torgerson T, Skinner M, et al. An assessment of transparency and reproducibility-related research practices in otolaryngology. *The Laryngoscope* 2020;130(8):1894-901. doi: 10.1002/lary.28322 [published Online First: 2019/10/09]
28. Statham EE, White SA, Sonwane B, et al. Primed to comply: Individual participant data sharing statements on ClinicalTrials.gov. *PloS one* 2020;15(2):e0226143. doi: 10.1371/journal.pone.0226143 [published Online First: 2020/02/19]
29. Mayer CS, Williams N, Gold S, et al. Sharing of Individual Participant Data from Clinical Trials: General Comparison and HIV Use Case. *AMIA Annual Symposium proceedings AMIA Symposium* 2019;2019:647-54. [published Online First: 2020/04/21]
30. Ali J, Califf R, Sugarman J. Anticipated Ethics and Regulatory Challenges in PCORnet: The National Patient-Centered Clinical Research Network. *Accountability in research* 2016;23(2):79-96. doi: 10.1080/08989621.2015.1023951 [published Online First: 2015/07/21]
31. Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15(2):125-34. doi: 10.1111/1467-8519.00220 [published Online First: 2001/11/08]
32. Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *Journal of clinical epidemiology* 2019;106:60-69. doi: 10.1016/j.jclinepi.2018.10.006 [published Online First: 2018/10/22]
33. Godolphin PJ, Bath PM, Montgomery AA. Short email with attachment versus long email without attachment when contacting authors to request unpublished data for a systematic review: a nested randomised trial. *BMJ open* 2019;9(1):e025273. doi: 10.1136/bmjopen-2018-025273 [published Online First: 2019/02/02]
34. Howe N, Giles E, Newbury-Birch D, et al. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *Journal of health services research & policy* 2018;23(2):123-33. doi: 10.1177/1355819617751555 [published Online First: 2018/04/15]
35. Merson L, Phong TV, Nhan le NT, et al. Trust, Respect, and Reciprocity: Informing Culturally Appropriate Data-Sharing Practice in Vietnam. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):251-63. doi: 10.1177/1556264615592387 [published Online First: 2015/08/25]
36. Cheah PY, Tangseefa D, Somsaman A, et al. Perceived Benefits, Harms, and Views About How to Share Data Responsibly: A Qualitative Study of Experiences With and Attitudes Toward Data Sharing Among

Research Staff and Community Representatives in Thailand. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):278-89. doi: 10.1177/1556264615592388 [published Online First: 2015/08/25]
37. Broes S, Verbaanderd C, Casteels M, et al. Sharing of Clinical Trial Data and Samples: The Cancer Patient Perspective. *Frontiers in medicine* 2020;7:33. doi: 10.3389/fmed.2020.00033 [published Online First: 2020/03/03]
38. Jones CW, Roberts BW, Platts-Mills TF. Patient Perspectives on Sharing Deidentified Trial Data. *Annals of internal medicine* 2016;165(10):748-49. doi: 10.7326/l16-0089 [published Online First: 2016/11/15]
39. Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *The New England journal of medicine* 2018;378(23):2202-11. doi: 10.1056/NEJMsa1713258 [published Online First: 2018/06/07]
40. Colombo C, Roberto A, Krleza-Jeric K, et al. Sharing individual participant data from clinical studies: a cross-sectional online survey among Italian patient and citizen groups. *BMJ open* 2019;9(2):e024863. doi: 10.1136/bmjopen-2018-024863 [published Online First: 2019/02/21]
41. Spence O, Onwuchekwa Uba R, Shin S, et al. Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials* 2018;19(1):269. doi: 10.1186/s13063-018-2651-2 [published Online First: 2018/05/05]
42. Krleža-Jerić K, Sim I, Marušić A, Reveiz L & Granados C. Medical Journal Editor Perspectives on Sharing Results Data From Patient-Oriented Research: A WAME Survey. 7th Peer Review Congress; 2009; Vancouver.
43. Chapman SJ, Shelton B, Mahmood H, Bhangu A, Fitzgerald JEF, Harrison EM. Promoting transparency in clinical research: Systematic review of disclosure and data-sharing policies in surgical journals. ASiT; 2014.
44. Chikramane A, Venkatraman L, Khanna V, Sinha S. Data sharing trends and integrity in the era of open science: a qualitative analysis of journals in top-ranked therapeutic areas. 13th Annual Meeting of ISMPP; 2017.
45. Vidal-Infer A, Tarazona B, Alonso-Arroyo A, et al. Public availability of research data in dentistry journals indexed in Journal Citation Reports. *Clinical oral investigations* 2018;22(1):275-80. doi: 10.1007/s00784-017-2108-0 [published Online First: 2017/03/28]
46. Vassar M, Jellison S, Wendelbo H, et al. Data sharing practices in randomized trials of addiction interventions. *Addictive behaviors* 2020;102:106193. doi: 10.1016/j.addbeh.2019.106193 [published Online First: 2019/11/27]
47. Nutu D, Gentili C, Naudet F, et al. Open science practices in clinical psychology journals: An audit study. *Journal of abnormal psychology* 2019;128(6):510-16. doi: 10.1037/abn0000414 [published Online First: 2019/08/02]
48. Gorman DM. Use of publication procedures to improve research integrity by addiction journals. *Addiction (Abingdon, England)* 2019;114(8):1478-86. doi: 10.1111/add.14604 [published Online First: 2019/03/10]
49. Gorman DM, Ferdinand AO. High impact nutrition and dietetics journals' use of publication procedures to increase research transparency. *Research Integrity and Peer Review* 2020;5(1):12. doi: 10.1186/s41073-020-00098-9
50. Almaqrami BS, Hua F, Liu Y, et al. Research waste-related editorial policies of leading dental journals: Situation 2018. *Oral diseases* 2020;26(3):696-706. doi: 10.1111/odi.13257 [published Online First: 2019/12/18]
51. Siebert M, Gaba JF, Caquelin L, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. *BMJ open* 2020;10(5):e038887. doi: 10.1136/bmjopen-2020-038887 [published Online First: 2020/06/01]
52. Aleixandre-Benavent R, Sapena AF, Ferrer SC, et al. Policies regarding public availability of published research data in pediatrics journals. *Scientometrics* 2019;118(2):439-51. doi: 10.1007/s11192-018-2978-1
53. DeVito NJ, French L, Goldacre B. Noncommercial Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *Jama* 2018;319(16):1721-23. doi: 10.1001/jama.2018.2841 [published Online First: 2018/05/02]
54. Whitlock EP, Dunham KM, DiGioia K, et al. Noncommercial US Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *JAMA network open* 2019;2(1):e187498. doi: 10.1001/jamanetworkopen.2018.7498 [published Online First: 2019/01/27]

55. Rollando P, Parc C, Naudet F, et al. [Data sharing policies of clinical trials funders in France]. *Therapie* 2020;75(6):527-36. doi: 10.1016/j.therap.2020.04.001 [published Online First: 2020/05/25]
56. Goldacre B, Lane S, Mahtani KR, et al. Pharmaceutical companies' policies on access to trial data, results, and methods: audit study. *BMJ (Clinical research ed)* 2017;358:j3334. doi: 10.1136/bmj.j3334 [published Online First: 2017/07/28]
57. Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. *BMC medicine* 2018;16(1):165. doi: 10.1186/s12916-018-1154-z [published Online First: 2018/09/29]
58. Gaba JF, Siebert M, Dupuy A, et al. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PloS one* 2020;15(8):e0237464. doi: 10.1371/journal.pone.0237464 [published Online First: 2020/08/21]
59. Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *The New England journal of medicine* 2005;352(21):2202-10. doi: 10.1056/NEJMsa044115 [published Online First: 2005/05/27]
60. Bosserdt M, Hamm B, Dewey M. Clinical trials in radiology and data sharing: results from a survey of the European Society of Radiology (ESR) research committee. *European radiology* 2019;29(9):4794-802. doi: 10.1007/s00330-019-06105-y [published Online First: 2019/02/28]
61. Rowhani-Farid A, Aldcroft A, Barnett AG. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Royal Society open science* 2020;7(3):191818. doi: 10.1098/rsos.191818 [published Online First: 2020/04/10]
62. Veroniki AA, Ashoor HM, Le SPC, et al. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *Journal of clinical epidemiology* 2019;113:176-88. doi: 10.1016/j.jclinepi.2019.05.031 [published Online First: 2019/06/04]
63. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ (Clinical research ed)* 2012;344:d7762. doi: 10.1136/bmj.d7762 [published Online First: 2012/01/05]
64. Nevitt SJ, Marson AG, Davie B, et al. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ (Clinical research ed)* 2017;357:j1390. doi: 10.1136/bmj.j1390 [published Online First: 2017/04/07]
65. Geifman N, Bollyky J, Bhattacharya S, et al. Opening clinical trial data: are the voluntary data-sharing portals enough? *BMC medicine* 2015;13:280. doi: 10.1186/s12916-015-0525-y [published Online First: 2015/11/13]
66. Kochhar S, Knoppers B, Gamble C, et al. Clinical trial data sharing: here's the challenge. *BMJ open* 2019;9(8):e032334. doi: 10.1136/bmjopen-2019-032334 [published Online First: 2019/08/24]
67. Schmidt H, Caldwell B, Mollet A, et al. An Industry Experience with Data Sharing. *The New England journal of medicine* 2018;379(21):2081-82. doi: 10.1056/NEJMc1805610 [published Online First: 2018/11/22]
68. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Scientific data* 2018;5:180268. doi: 10.1038/sdata.2018.268 [published Online First: 2018/11/28]
69. So D, Knoppers BM. Ethics approval in applications for open-access clinical trial data: An analysis of researcher statements to clinicalstudydatarequest.com. *PloS one* 2017;12(9):e0184491. doi: 10.1371/journal.pone.0184491 [published Online First: 2017/09/09]
70. Boutron I, Dechartres A, Baron G, et al. Sharing of Data From Industry-Funded Registered Clinical Trials. *Jama* 2016;315(24):2729-30. doi: 10.1001/jama.2016.6310 [published Online First: 2016/07/02]
71. Miller J, Ross JS, Wilenzick M, et al. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *BMJ (Clinical research ed)* 2019;366:l4217. doi: 10.1136/bmj.l4217 [published Online First: 2019/07/12]
72. Coady SA, Mensah GA, Wagner EL, et al. Use of the National Heart, Lung, and Blood Institute Data Repository. *The New England journal of medicine* 2017;376(19):1849-58. doi: 10.1056/NEJMsa1603542 [published Online First: 2017/04/13]
73. Shmueli-Blumberg D, Hu L, Allen C, et al. The national drug abuse treatment clinical trials network data share project: website design, usage, challenges, and future directions. *Clinical trials (London, England)* 2013;10(6):977-86. doi: 10.1177/1740774513503522 [published Online First: 2013/10/03]
74. Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. *Clinical trials (London, England)* 2018;15(4):413-23. doi: 10.1177/1740774518769655 [published Online First: 2018/04/21]

75. Bertagnolli MM, Sartor O, Chabner BA, et al. Advantages of a Truly Open-Access Data-Sharing Model. *The New England journal of medicine* 2017;376(12):1178-81. doi: 10.1056/NEJMsb1702054 [published Online First: 2017/03/23]
76. Asare AL, Carey VJ, Rotrosen D, et al. Clinical trial data access: Opening doors with TrialShare. *The Journal of allergy and clinical immunology* 2016;138(3):724-26. doi: 10.1016/j.jaci.2016.05.034 [published Online First: 2016/07/31]
77. MRC. MRC Clinical Trials Transparency Review 2017 [Available from: https://mrc.ukri.org/documents/pdf/clinical-trials-review-2017/ accessed 2021/07/01.
78. Zhu CS, Pinsky PF, Moler JE, et al. Data sharing in clinical trials: An experience with two large cancer screening trials. *PLoS medicine* 2017;14(5):e1002304. doi: 10.1371/journal.pmed.1002304 [published Online First: 2017/05/26]
79. Chalmers I, Altman DG, McHaffie H, et al. Data sharing among data monitoring committees and responsibilities to patients and science. *Trials* 2013;14:102. doi: 10.1186/1745-6215-14-102 [published Online First: 2013/06/21]
80. Tudur Smith C, Nevitt S, Appelbe D, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials* 2017;18(1):319. doi: 10.1186/s13063-017-2067-4 [published Online First: 2017/07/18]
81. Strom BL, Buyse M, Hughes J, et al. Data sharing, year 1--access to data from industry-sponsored clinical trials. *The New England journal of medicine* 2014;371(22):2052-4. doi: 10.1056/NEJMp1411794 [published Online First: 2014/10/16]
82. Strom BL, Buyse ME, Hughes J, et al. Data Sharing - Is the Juice Worth the Squeeze? *The New England journal of medicine* 2016;375(17):1608-09. doi: 10.1056/NEJMp1610336 [published Online First: 2016/10/27]
83. Ross JS, Ritchie JD, Bamford S, Berlin JA, Karla Childers K, Desai N, Gamble GM, Gross CP, Richard S, Lehman RS, Lins P, Morris S, Waldstreicher J, Krumholz H. Sharing Data Through the Yale University Open Data Access (YODA) Project: Early Experience. 8th Peer Review Congress; 2017; Chicago.
84. Lovato L. Exploring the potential of data sharing: The SPRINT Challenge in Abstracts from the SCT 39th Annual Meeting (2018). *Clinical trials (London, England)* 2018;15(2_suppl):S35-s192. doi: 10.1177/1740774518790846 [published Online First: 2018/08/22]
85. Kawahara T, Fukuda M, Oba K, et al. Meta-analysis of randomized clinical trials in the era of individual patient data sharing. *International journal of clinical oncology* 2018;23(3):403-09. doi: 10.1007/s10147-018-1237-z [published Online First: 2018/01/14]
86. Ebrahim S, Sohani ZN, Montoya L, et al. Reanalyses of randomized clinical trial data. *Jama* 2014;312(10):1024-32. doi: 10.1001/jama.2014.9646 [published Online First: 2014/09/10]
87. Naudet F, Sakarovitch C, Janiaud P, 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 The BMJ and PLOS Medicine. *BMJ (Clinical research ed)* 2018;360:k400. doi: 10.1136/bmj.k400 [published Online First: 2018/02/15]
88. Ross JS, Ritchie JD, Finn E, et al. Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ open* 2016;6(9):e012769. doi: 10.1136/bmjopen-2016-012769 [published Online First: 2016/09/28]
89. Vaduganathan M, Nagarur A, Qamar A, et al. Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation* 2018;137(9):938-47. doi: 10.1161/circulationaha.117.031883 [published Online First: 2017/11/15]
90. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PloS one* 2007;2(3):e308. doi: 10.1371/journal.pone.0000308 [published Online First: 2007/03/22]
91. ECRIN. ECRIN metadata repository [Available from: https://ecrin.org/tools/clinical-research-metadata-repository accessed 2021/07/01.
92. Doshi P, Shamseer L, Jones M, et al. Restoring biomedical literature with RIAT. *Bmj* 2018;361:k1742. doi: 10.1136/bmj.k1742 [published Online First: 2018/04/28]
93. Takata J, Sondo P, Humphreys GS, et al. The WorldWide Antimalarial Resistance Network Clinical Trials Publication Library: A Live, Open-Access Database of Plasmodium Treatment Efficacy Trials. *The American journal of tropical medicine and hygiene* 2020;103(1):359-68. doi: 10.4269/ajtmh.19-0706 [published Online First: 2020/05/21]
94. Crossmark [Available from: https://www.crossref.org/services/crossmark/ accessed 2021/07/01.
95. Devereaux PJ, Guyatt G, Gerstein H, et al. Toward Fairness in Data Sharing. *The New England journal of medicine* 2016;375(5):405-7. doi: 10.1056/NEJMp1605654 [published Online First: 2016/08/16]

96. Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic reviews* 2012;1:2. doi: 10.1186/2046-4053-1-2 [published Online First: 2012/05/17]

**ABBREVIATIONS**

ANR: Agence Nationale de la Recherche
ASAPbio: Accelerating Science and Publication in biology
BioLINCC: Biological Specimen and Data Repository Information Coordination Center
CDAS: Cancer Data Access System
CIHR: Canadian Institutes of Health Research
CSDR: Clinical Study Data Request
CTN: Clinical Trials Network
DFG: Deutsche Forschungsgemeinschaft
DGOS : Direction Générale de l'Offre de Soins
Drum: Data Repository for University of Minnesota
EBCTG: Early Breast Cancer Trialists' Collaborative Group
EC Europe: European Commission
EFPIA: European Federation of Pharmaceutical Industries and Associations
F1000Research: Faculty of 1000 Research
FAIR: Findable, Accessible, Interoperable and Reusable
ICMJE: International Committee of Medical Journal Editors
ICPSR: Inter-university Consortium for Political and Social Research
IOM: Institute Of Medicine
IPD: Individual Participant Data
ITN Trialshare: Immune Tolerance Network TrialShare
MMMP: Melanoma Molecular Map Project
MRC UK: Medical Research Council
NHMRC: National Health and Medical Research Council
NIDA: National Institute on Drug Abuse
NIDDK: National institute of Diabetes and Digestive and Kidney Diseases
NIH: National Institute of Health
NIH BioLINCC: National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center
NIHR: National Institute of Health Research
NIMH NDCT: National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness
NSFC: National Natural Science Foundation of China
PCORNeT: The National Patient-centered Clinical research Network
PHRC: Le programme hospitalier de recherche clinique
PHRMA: Pharmaceutical Research and Manufacturers of America
PI: Principal Investigator
PLOS: Public Library Of Science
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: extension for Scoping Reviews
ProAct: Pooled Resource Open Access Clinical trials database
RCT: Randomised clinical trial
RDA: Research Data Alliance
SOAR: the Supporting Open Access for Researchers initiative
SND: Swedish National Data Service
TBI-IMPACT: Traumatic Brain Injury– International Mission for Prognosis and Analysis of Clinical trials in Traumatic brain injury
The BMJ: The British Medical Journal
UK: United Kingdom
UKCRC: UK Clinical Research Collaboration
UMIN: University Medical Hospital Information Network
US: United States of America
US DoD: United States Department of Defense
Vivli: adapted from the Greek "Vivliothiki" (library) and the Latin root "viv" (life)
WWARN: World Wide Antimalarial Resistance Network
YODA: the Yale University Open Data Access Project

**FIGURES**

**Figure 1**: PRISMA flow diagram (PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses).
* For National Institute of Health (NIH US), the answer we received was not informative

**Figure 2**: Proportion of the 93 references exploring each outcome domain
Study designs considered:
- Experimental: prospective research that implies testing the impact a strategy (e.g. randomised controlled trial)
- Survey: a general overview, exploration, or description of individuals and/or research objects;
- Metrics: descriptive metrics from each initiative provided by the initiative;
- Qualitative: research that relies on non-numerical data to understand concepts, opinions or experiences.
- Other: any other research not covered above (e.g. case studies, environmental scans, etc.)"

**Figure 3**: Outcomes used to assess current data-sharing practices for individual patient data for clinical trials organized per outcome domain and number of studies exploring these outcomes.
Study designs considered:
- Experimental: prospective research that implies testing the impact a strategy (e.g. randomised controlled trial)
- Survey: a general view, exploration, or description of individuals and/or research objects;
- Metrics: descriptive metrics from each initiative provided by the initiative;
- Qualitative: research that relies on non-numerical data to understand concepts, opinions or experiences.
- Other: any other research not covered above (e.g. case studies, environmental scans, etc.)"

**Figure 4**: Intent to share

Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference

a: The proportion is 73 % if the purpose is a re-analysis

b: 54 participants out of 60 had an opinion about data-sharing (the others had no knowledge or no opinion)

c: An additional 25 % were undecided

d: The proportion is 19 % for requiring a data-sharing plan

e: 35 % have a data-sharing policy (encouraging data-sharing)

f: Only 2 with a mandatory policy

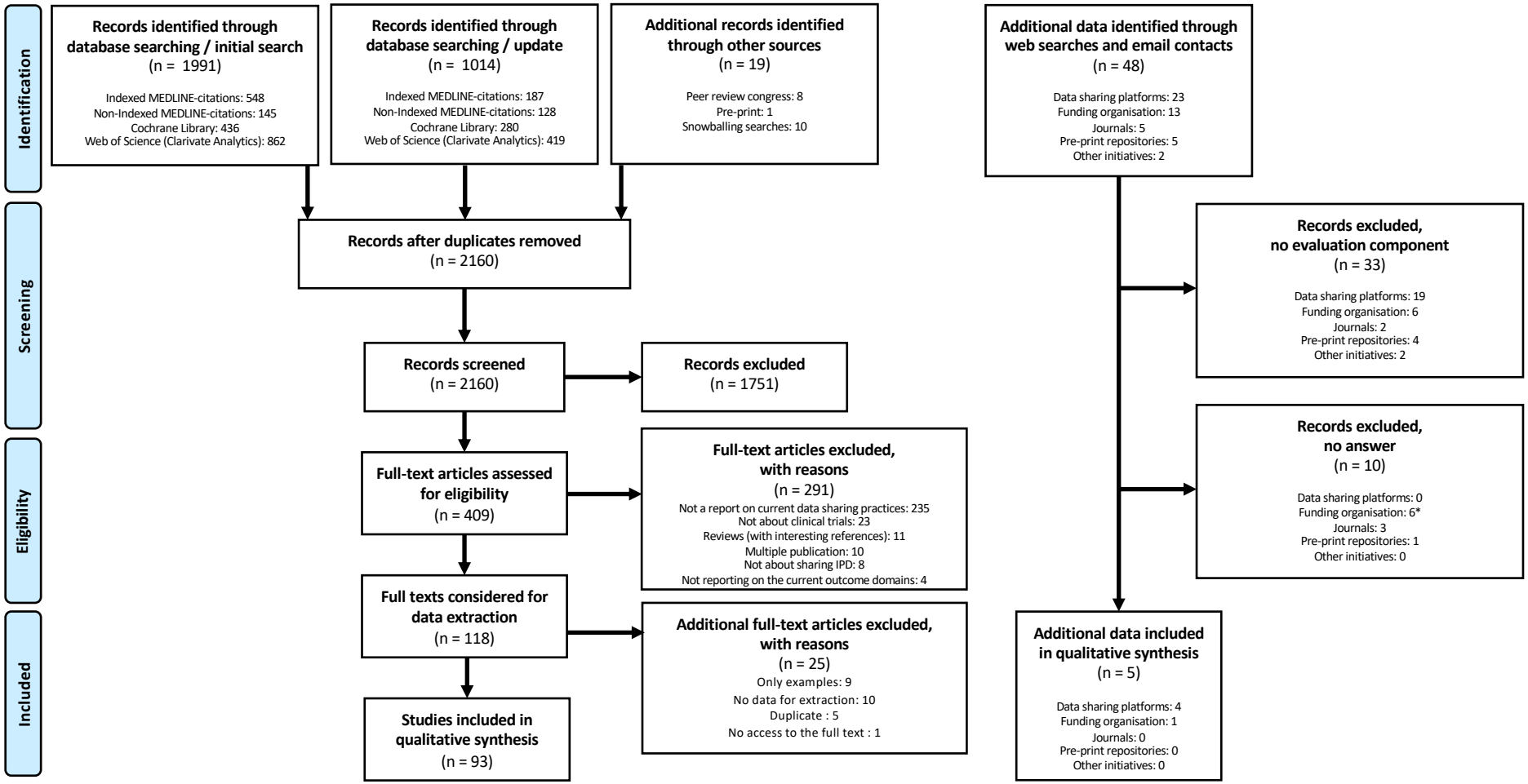
g: The proportion is 71% for a sample of all companies (not only the top 25)

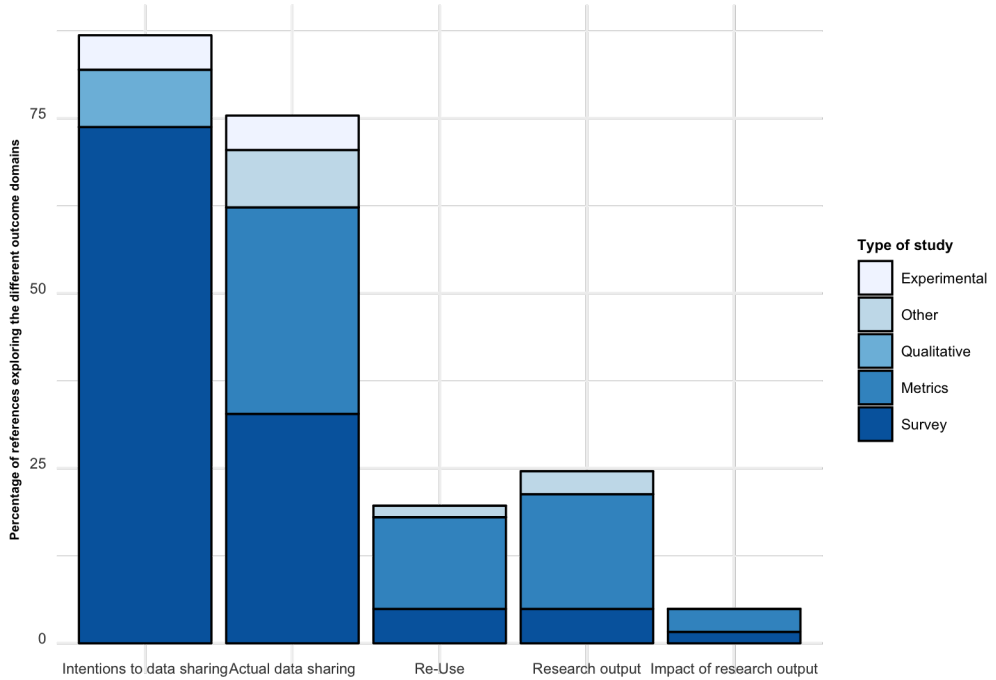
In DeVito et al. 2018, we extracted the information on policies that made data-sharing mandatory (i.e. a requirement to share the data).

**Figure 5**: Actual data-sharing

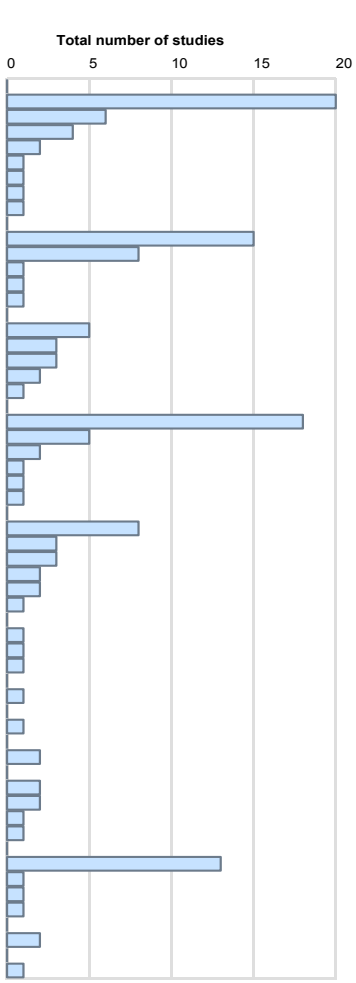
Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference

**Figure 6**: Temporal trends, number and type of published output from CSDR and YODA
- Blue: YODA
- Red: CSDR

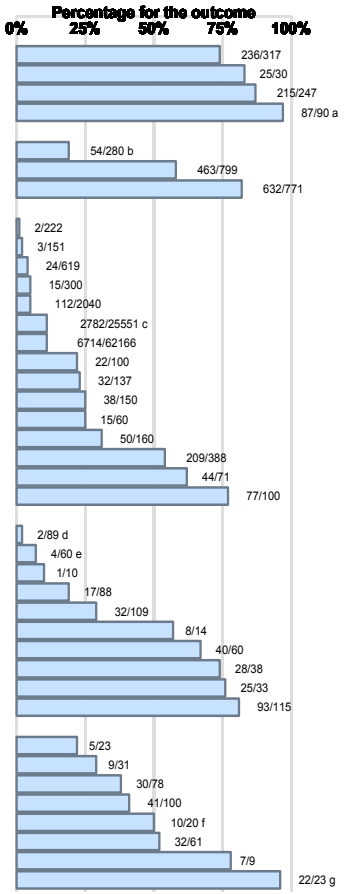

Identification
Screening
Eligibility
Included
Records identified through database searching / initial search (n = 1991)
Indexed MEDLINE-citations: 548
Non-Indexed MEDLINE-citations: 145
Cochrane Library: 436
Web of Science (Clarivate Analytics): 862
Records identified through database searching / update (n = 1014)
Indexed MEDLINE-citations: 187
Non-Indexed MEDLINE-citations: 128
Cochrane Library: 280
Web of Science (Clarivate Analytics): 419
Additional records identified through other sources (n = 19)
Peer review congress: 8
Pre-print: 1
Snowballing searches: 10
Additional data identified through web searches and email contacts (n = 48)
Data sharing platforms: 23
Funding organisation: 13
Journals: 5
Pre-print repositories: 5
Other initiatives: 2
Records after duplicates removed (n = 2160)
Records screened (n = 2160)
Records excluded (n = 1751)
Records excluded, no evaluation component (n = 33)
Data sharing platforms: 19
Funding organisation: 6
Journals: 2
Pre-print repositories: 4
Other initiatives: 2
Full-text articles assessed for eligibility (n = 409)
Full-text articles excluded, with reasons (n = 291)
Not a report on current data sharing practices: 235
Not about clinical trials: 23
Reviews (with interesting references): 11
Multiple publication: 10
Not about sharing IPD: 8
Not reporting on the current outcome domains: 4
Records excluded, no answer (n = 10)
Data sharing platforms: 0
Funding organisation: 6*
Journals: 3
Pre-print repositories: 1
Other initiatives: 0
Full texts considered for data extraction (n = 118)
Additional full-text articles excluded, with reasons (n = 25)
Only examples: 9
No data for extraction: 10
Duplicate : 5
No access to the full text : 1
Additional data included in qualitative synthesis (n = 5)
Data sharing platforms: 4
Funding organisation: 1
Journals: 0
Pre-print repositories: 0
Other initiatives: 0
Studies included in qualitative synthesis (n = 93)


Percentage of references exploring the different outcome domains
75
50
25
0
Intentions to data sharing
Actual data sharing
Re-Use
Research output
Impact of research output
Type of study
Experimental
Other
Qualitative
Metrics
Survey

| Outcome domains | Outcomes | Experimental | Survey | Metrics | Qualitative | Other |
|---|---|---|---|---|---|---|
| Intentions to data sharing | **For trialists** | . | . | . | . | . |
| | Intentions to share data | 2 | 18 | . | . | . |
| | Existence of a data sharing plan/data sharing statement | . | 6 | . | . | . |
| | Type of data sharing plan | . | 4 | . | . | . |
| | Support for data sharing | . | 2 | . | . | . |
| | Willingness to store IPD in a central repository, conditions for storage | . | 1 | . | . | . |
| | Most pressing ethical issues | . | 1 | . | . | . |
| | Response to a request for data sharing and time to response | 1 | . | . | . | . |
| | Preparedness for DS | 1 | . | . | . | . |
| | **For publishers/funders** | . | . | . | . | . |
| | Existence of a data sharing policy/intent to share | . | 15 | . | . | . |
| | Type of data sharing policy | . | 8 | . | . | . |
| | Journals requiring participant level data | . | 1 | . | . | . |
| | Prohibiting data sharing in agreements with industry | . | 1 | . | . | . |
| | Existence of a data sharing policy for clinical trial units | . | 1 | . | . | . |
| | **For trial participants** | . | . | . | . | . |
| | Willingness to share data | . | 4 | . | 1 | . |
| | Barriers to share data | . | 2 | . | 1 | . |
| | Patient Opinions on the Release of Deidentified Individual-Patient Data | . | 3 | . | . | . |
| | Views, experience and attitudes towards DS | . | . | . | 2 | . |
| | Information in consent forms | . | 1 | . | . | . |
| Actual data sharing | **For re-users (e.g. IPD meta-analysts)** | . | . | . | . | . |
| | Data availability (yes/no) | 3 | 13 | . | . | 2 |
| | Data release after a request | . | 4 | 1 | . | . |
| | Data availability (time) | 1 | . | 1 | . | . |
| | Data availability (completeness) | 1 | . | . | . | . |
| | Rate of IPD meta-analyses / all requests | . | . | 1 | . | . |
| | Reasons for request, research plan and experience of DS | . | 1 | . | . | . |
| | **For repositories/platforms** | . | . | . | . | . |
| | Data release after a request | . | 2 | 5 | . | 1 |
| | Approval of data release | . | 2 | 1 | . | . |
| | Data availability | . | 3 | . | . | . |
| | Number of downloads | . | . | 2 | . | . |
| | Number of DS agreement, speed of data availability | . | 1 | 1 | . | . |
| | Database access | . | . | 1 | . | . |
| | **For other stakeholders** | . | . | . | . | . |
| | Number of downloads and page hits for any public access | . | . | 1 | . | . |
| | Re-use of data by independent data monitoring committees | . | . | . | . | 1 |
| | Cost of preparing the data for data generators | . | . | . | . | 1 |
| Re-use | **Further additional analyses** | . | . | . | . | . |
| | Approved proposals by subject areas | . | . | 1 | . | . |
| | **IPD meta-analyses** | . | . | . | . | . |
| | Proposals for IPD-meta-analyses and one concrete example | . | . | 1 | . | . |
| | **Re-analyses** | . | . | . | . | . |
| | Reproducibility on primary outcomes | . | 2 | . | . | . |
| | **Any type of re-use** | . | . | . | . | . |
| | Type of re-use | . | . | 2 | . | . |
| | Progression of the analysis | . | . | 1 | . | 1 |
| | Listing of type of studies performed | . | . | 1 | . | . |
| | Project completion | . | 1 | . | . | . |
| Output from data sharing | **Published re-use** | . | . | . | . | . |
| | Publication of re-use (papers) | . | 2 | 9 | . | 2 |
| | Manuscripts in peer review | . | . | 1 | . | . |
| | Request discontinued | . | . | 1 | . | . |
| | Communication of re-use (oral presentation, posters) | . | . | . | . | 1 |
| Impact of research output | **Quantitative metrics for published articles** | . | . | . | . | . |
| | Citation metrics | . | 1 | 1 | . | . |
| | **Identification of a new finding** | . | . | . | . | . |
| | Crowd voting | . | 1 | . | . | . |



| Reference | Type | Population | Time point |
|---|---|---|---|
| **A: Intent to share: surveys of trialists** | | | |
| Rathi, 2012 | Overall | Trials published in 6 high impact journals | 2010-2011 |
| Tudur-Smith, 2014 | Overall (central storage of their IPD) | Reviewers of Cochrane IPD meta-analysis group | 2011 |
| Yuanyuan, 2017 | Overall (endorsment of DS) | Trials publihed in Chinese Medical Journal | 2016 |
| Tannenbaum, 2018 | For an IPD meta-analysis | Trials published in 3 high impact journals with data sharing policies | 2012-2016 |
| **B: Intent to share: surveys of trial participants** | | | |
| Colombo, 2017 | Having knowlegde about and being in favor of data sharing | Italian patients and citizen groups | 2017 |
| Jones, 2016 | Favor or strongly favor data sharing | Patients in a Usemergency department | 2015 |
| Mello, 2018 | Perception that the benefits of data sharing outweighed the negative aspects | Patients from 3 US medical centers | Unclear |
| **C: Intent to share: data sharing statements** | | | |
| Kemper, 2020 | Overall | Reproductive endocrinology and infertility articles (study mix) | 2013, 2018 |
| Johnson, 2020 | Overall | 300 otolaryngology research studies | 2014-2018 |
| Gabelica, 2019 | Overall | RCTs in 7 high-ranked anesthesiology journals | 2014-2016 |
| Papageorgios, 2019 | Open data | Trials in orthodontics and periodontics | 2017-2018 |
| Statham, 2020 | Overall | CT.gov | 2018 |
| Bergeris, 2018 | Overall | CT.gov | Up to august 2017 |
| Mayer, 2019 | Studies with a data sharing plan | CT.gov | 2015-2018 |
| Siebert, 2020b | Overall | ICMJE affiliates (after policy) | 2019 |
| Kaufman, 2019 | Overall | RCTs in 11 selected journals (before policy) | 2018 |
| Kaufman, 2019b | Overall | RCTs in 11 selected journals (after policy) | 2018 |
| Murugiah, 2016 | Data made available | Clinical trials (> 5000 patients), from clinicaltrials.gov | Up to 2015 |
| Rowhani-Farid, 2016 | Overall | The BMJ (all research papers) | 2009-2015 |
| Griswold M, 2013 | Overall | Ann. Int. Med. (all research papers) | 2008-2012 |
| Laine, 2009 | Overall | Ann. Int. Med. (all research papers) | 2008 |
| Siebert, 2020 | Overall | ICMJE members (after policy) | 2019 |
| **D: Intent to share: journal data sharing policies** | | | |
| Krleza-Jeric, 2009 | Require IPD | Members of World Association of Medical Editors | 2009 |
| Chickramane, 2017 | Require IPD | 15 oncology, 15 central nervous system, 15 cardiology/endocrinology and 15 respiratory journals | Unknown |
| Chapman, 2014 | Data sharing policy | High impact surgical journals | 2009-2012 |
| Vidal-Infer, 2018 | Accept IPD as a complementary material | Dental journals | 2014 |
| Almaqrami, 2020 | Data sharing policy | Dental journals | 2018 |
| Vasar, 2020 | Data sharing policy | 15 high-impact addiction journals | 2013-2018 |
| Nutu, 2019 | Data sharing policy | Clinical psychology journals | 2017 |
| Gorman, 2019 | Data sharing policy | High impact addiction journals | 2018 |
| Gorman, 2020 | Data sharing policy | Nutrition and dietetics journals | 2018 |
| Alexandre-Benavent, 2019 | Data sharing policy | Paediatric journals | 2012-2016 |
| **E: Intent to share: funders and clinical trial units** | | | |
| Hopkins, 2016 | Data sharing policy | UK CTUs | 2014 |
| Rollando, 2020 | Data sharing policy | Clinical trial funders in France | 2019 |
| Gaba, 2020 | Data sharing policy | Non-commercial | 2016-2018 |
| Gaba, 2020b | Data sharing policy | Commercial funders | 2016-2018 |
| de Vito, 2018 | Data sharing requirement | Top non commercial funders | 2017 |
| Hopkins, 2018 | Data sharing policy | Clinical trials sponsored by the pharmaceutical industry published in the top 10 medical journals | 2015 |
| Whitlock, 2019 | Data sharing policy | Non-commercial funders in the US | 2018 |
| Goldacre, 2017 | Data sharing policy | Top 25 pharmaceutical companies by revenue | 2016 |



**A: Actual data sharing by re-users: surveys related to published studies**

| Reference | Data source | Sample selection | Time period |
|---|---|---|---|
| Vassar, 2020 | 15 high ranked addiction journals | Consecutive | 2013-2018 |
| Gabelica, 2019 | 7 high ranked addiction journals | Consecutive | 2014-2016 |
| Savage, 2009 | PLOS Medicine, PLOS clinical trials | Unclear | 2009 |
| Yuanyuan, 2017 | Chinese Medical Journal | Consecutive | 2016 |
| Hopkins, 2018 | Top 10 general and internal medical journals | Consecutive, industry-sponsored trials | 2015 |
| Rowhani-Farid, 2016 | BMJ | Random, subsample RCTs | 2009-2015 |
| Rathi, 2012 | 6 general medical journals | Consecutive | 2010-2011 |
| Naudet, 2018 | BMJ, PLOS Medicine | Consecutive, RCTs | 2013-2016 |
| Tannenbaum, 2018 | BMJ, PLOS Medicine, Ann. Inn. Med. | Selected, partly restricted to authors willing to share data | 2012-2016 |

**B: Actual data sharing by re-users: data related to IPD meta-analyses**

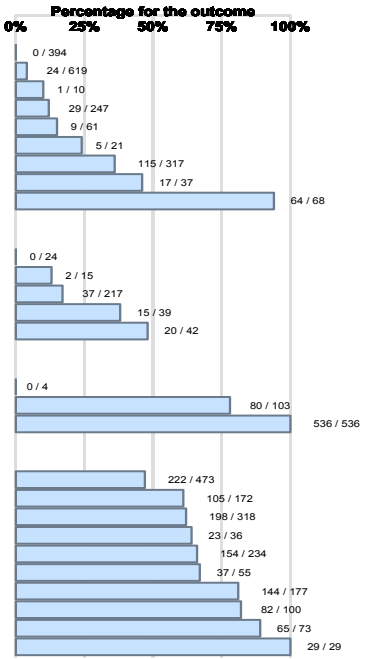
| Reference | Number of studies with IPD sharing | Time point | Comment |
|---|---|---|---|
| Mayo-Wilson, 2015 | 0/24 (0%) | Contacted in 2014 | Commercial sponsor, trials with one medicinal product |
| Kawahara, 2018 | 2/15 (13%) through CSDR | Unknown | 13 requested from authors directly, project in progress |
| Villein, 2015 | 37/217 (17%) | RCTs with results published since 2000 | |
| Nevitt,2017 | 15/35 (38%), 4/15 through CSDR | End 2015 | |
| Hee, 2016 | 20/42 (48%) | RCTs until 2011 | |

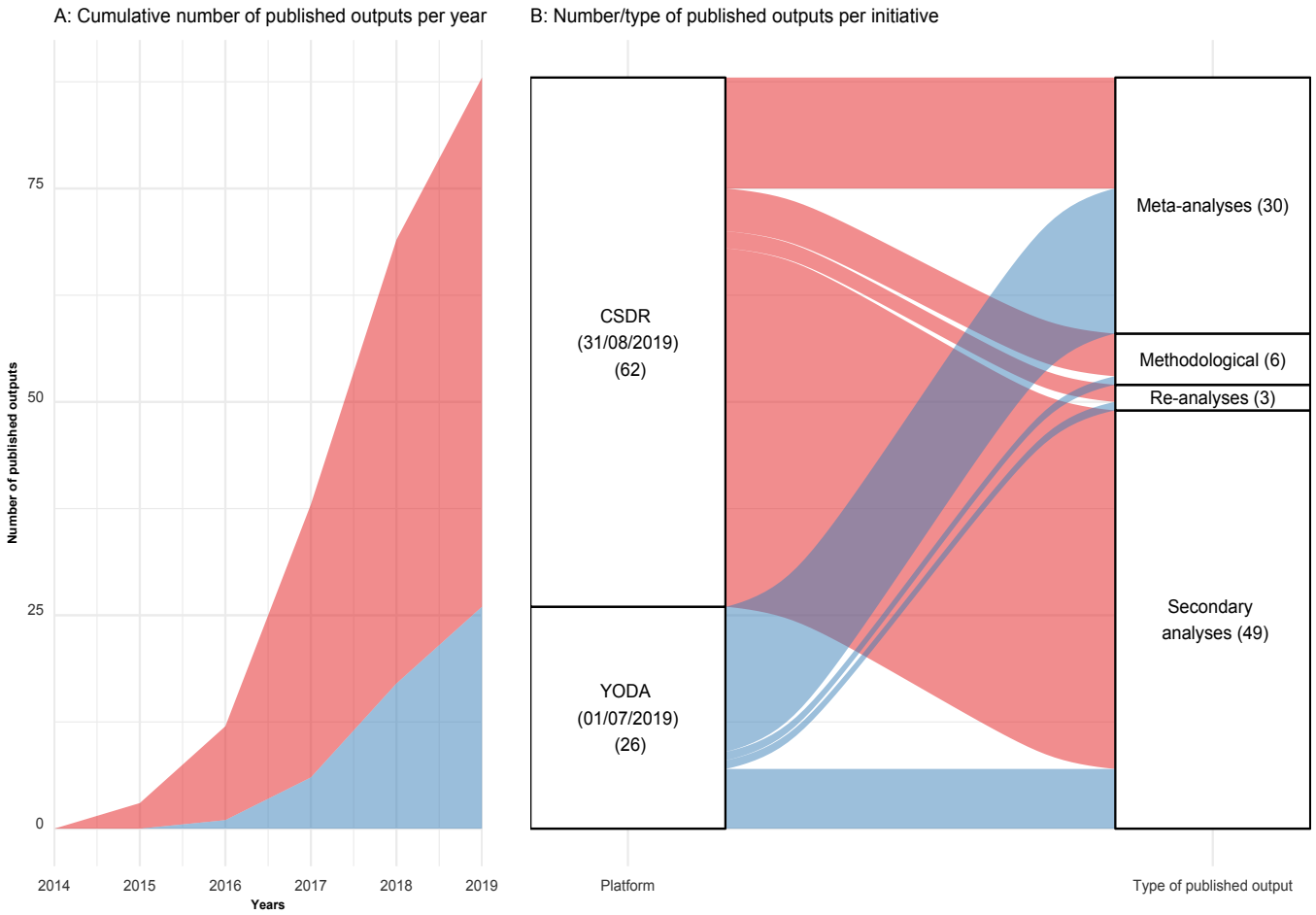
**C: Actual data sharing by re-users: repositories/platforms from the viewpoint of the user**

| Reference | Repository_platform | Time point of submission | Comment |
|---|---|---|---|
| Geifman, 2015 | CSDR | 12/2014-1/2015 | |
| Sydes, 2015 | MRC CTU | 2012-2014 | 103 requests to 54 trials |
| Ross, 2016 | BioLINCC | 2007-2014 | Survey of investigators who received access to BioLINCC |

**D: Actual data sharing by re-users: survey of repositories/platforms**

| Reference | Platform | Time point of assessment | Comment |
|---|---|---|---|
| Kochar, 2019 | CSDR | 2014-2019 | |
| So, 2017 | CSDR | February 2017 | |
| Vaduganathan, 2018 | CSDR | May 2017 | |
| Strom, 2014 | CSDR | May 2014 | |
| Navar, 2016 | CSDR, YODA, SOAR | December 2015 | |
| Schmidt, 2018 | CSDR | 2014-2017 | Boehringer-Ingelheim's studies |
| Strom, 2016 | CSDR | November 2015 | 177 for 237 trials |
| Ross, 2018 | YODA | August 2018 | |
| Ross, 2017 | YODA | June 2017 | 73 for 159 trials |
| Krumholz, 2016 | YODA | 2015 | |




A: Cumulative number of published outputs per year
Number of published outputs
0
25
50
75
2014
2015
2016
2017
2018
2019
Years
B: Number/type of published outputs per initiative
CSDR
(31/08/2019)
(62)
YODA
(01/07/2019)
(26)
Platform
Meta-analyses (30)
Methodological (6)
Re-analyses (3)
Secondary
analyses (49)
Type of published output

# Supplementary material 1: Information sources

For commercial sponsors, we considered:
1/ Clinical Study Data Request (CSDR),
2/ the Yale University Open Data Access Project (YODA),
3/ the Supporting Open Access for Researchers initiative (SOAR),
4/ ViVli.

For non-commercial sponsor, we considered:
1/ the National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness (NIMH NDCT),
2/ The National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center (NIH BioLINCC),
3/ B2Share,
4/ Dryad,
5/ the Data Repository for University of Minnesota (Drum),
6/ EASY,
7/ Edinburgh DataShare,
8/ FigShare,
9/ the Inter-university Consortium for Political and Social Research (ICPSR),
10/ the Swedish National data Service (SND),
11/ the University Medical Hospital Information Network (UMIN),
12/ Zenodo,
13/ the Early Breast Cancer Trialists' Collaborative Group (EBCTG),
14/ FreeBird,
15/ Traumatic Brain Injury – International Mission for Prognosis and Analysis of Clinical trials in TBI (TBI-IMPACT),
16/ Melanoma Molecular Map Project (MMMP),
17/ National institute of Diabetes and Digestive and Kidney Diseases (NIDDK),
18/ Immune Tolerance Network TrialShare (ITN Trialshare),
19/ Child Abuse,
20/ Pooled Resource Open Access Clinical trials database (ProAct).

For the different funders:
1/ National Institute of Health (NIH US),
2/ European Commission (EC Europe),
3/ Medical Research Council (MRC UK),
3/ Le programme hospitalier de recherche clinique (DGOS France),
4/ L'Agence nationale de la recherche (ANR France),
5/ Department of Defense (US DoD),
6/ Wellcome Trust UK,

7/ Canadian Institutes of Health Research (CIHR Canada),
8/ National Health and Medical Research Council (NHMRC Australia),
9/ Deutsche Forschungsgemeinschaft (DFG Germany),
10/ National Natural Science Foundation of China (NSFC China),
11/ National Institute of Health Research (NIHR UK),
12/ Gates Foundation US.

# Supplementary material 2: Literature searches

The initial algorithm for the literature search, detailed in the registered protocol (osf.io/pb8cj), was updated on October 29th 2018 to include broader search terms.

| **Name of Database** | **Host, search interface** | **Initial search** Search date 2018-10-29 | | **Update search** Search date 2020-09-11 Publication year 2018-2020 | |
|---|---|---|---|---|---|
| | | **Update status of the database** | **Results** | **Update status of the database** | **Results** Publication year From 2018-2020 |
| Medline | Wolters Kluwer / Ovid | 1946 to October Week 3 2018 | 548 | 1946 to September Week 1 2020 | 187 |
| Medline daily update | | October 25, 2018 | | September 09, 2020 | |
| MEDLINE In-Process & Other Non-Indexed Citations | | | 145 | 1946 to September 09, 2020 | 128 |
| MEDLINE Epub Ahead of Print | | | | September 09, 2020 | |
| Cochrane Library: Cochrane Reviews | Wiley | Issue 10 of 12, October 2018 | 19 | Issue 9 of 12, September 2020 | 12 |
| Cochrane Protocols | | | 1 | | 0 |
| Cochrane Central Register of Controlled Trials | | Issue 9 of 12, September 2018 | 416 | Issue 9 of 12, September 2020 | 268 |
| Science Citation Index | Clarivate Analytics / Web of Science | 1945 –present (2018-10-26) | 862 | 2018 –present (2020-09-10) | 419 |
| Social Science Citation Index | | 1956-present (2018-10-26) | | | |

| | | | | |
|---|---|---|---|---|
| **Total with duplicates** | | **1991** | | **1014** |
| **Total without duplicates** | | **1544** | | **763** |
| **New citations 2018-2020 without overlap from initial search** | | | | **597** |
| **Total without overlap initial search and update search (see PRISMA flow diagram)** | | **2141 (1544 + 597)** | | |

**MEDLINE Databases: Host: Wolters Kluwer, search interface: Ovid**

**1. Indexed MEDLINE-citations:**
Search Strategy:

| # | **Searches** | **Results Initial search** Search date: 2018-10-29: MEDLINE 1946 to October Week 3 2018, MEDLINE Daily Update October 25, 2018 | **Results Update search** Search date: 2020-09-11: MEDLINE 1946 to September Week 1 2020, MEDLINE (Daily Update September 09, 2020. | **Annotations** |
|---|---|---|---|---|
| 1 | exp Access to Information/ | 6845 | 7597 | Concept data sharing: MeSH terms |
| 2 | Information Dissemination/ | 14697 | 16894 | |
| 3 | exp *"Information Storage and Retrieval"/ | 52187 | 58658 | |
| 4 | data collection/ | 87165 | 89553 | |
| 5 | datasets as topic/ | 2259 | 4417 | |
| 6 | or/1-5 | 157717 | 170739 | |
| 7 | exp clinical trial/ | 809623 | 868410 | Concept clinical trials: MeSH terms or textwords |
| 8 | exp clinical trial as topic/ | 318580 | 345552 | |

| | | | | |
|---|---|---|---|---|
| 9 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 879338 | 997736 | |
| 10 | Meta-Analysis as Topic/ | 16485 | 18279 | Concept Meta-analysis: MeSH |
| 11 | meta-analysis/ | 93492 | 119228 | |
| 12 | or/7-11 | 1489253 | 1650537 | Concept Clinical Trials OR Meta-analysis |
| 13 | 6 and 12 | 10206 | 11264 | Combination of concepts: data sharing (MeSH only) AND (clinical trials OR meta-analysis) |
| 14 | (data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 7822 | 9921 | Concept data sharing: Textwords |
| 15 | 13 and 14 | 325 | 422 | data sharing (MeSH terms) AND data sharing (textwords) AND (clinical trials OR meta-analysis): 1. interim result |
| 16 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 756 | 993 | Concept sharing IPD (textwords) |
| 17 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 23 | 38 | |
| 18 | or/16-17 | 772 | 1017 | |
| 19 | 12 and 18 | 129 | 179 | (Clinical trials OR meta-analysis) AND textwords for sharing IPD: 2. interim result |
| 20 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3196 | 4262 | Concept data sharing textwords |
| 21 | 12 and 20 | 393 | 539 | (Clinical trials OR meta-analysis) AND textwords data sharing: 3. interim result |
| 22 | 15 or 19 or 21 | 557 | 738 | OR-combination of interim results |
| 23 | exp animals/ not humans/ | 4508403 | 4732433 | Exclusion of animals only |

| | | | | |
|---|---|---|---|---|
| 24 | 22 not 23 | 552 | 732 | |
| 25 | limit 24 to (english or french or german or spanish) | 548 | 725 | Restriction to English, German, French, Spanish: Final result for indexed Medline citations |
| | | | 187 | Update search: limit 25 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Exp **term*/ = MeSH as major topic incl. narrower terms as major topic
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters
adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**2. Non-Indexed MEDLINE-citations:**

| # | **Searches** | **Results Initial search** Search date: 2018-10-29: MEDLINE In-Process & Other Non-Indexed Citations October 25, 2018, MEDLINE Epub Ahead of Print October 25, 2018 | **Results Update search** Search date: 2020-09-11: MEDLINE Epub Ahead of Print September 09, 2020, MEDLINE In-Process & Other Non-Indexed Citations 1946 to September 09, 2020 | **Annotations** |
|---|---|---|---|---|
| 1 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 215 | 323 | Concept data sharing |

| | | | | |
|---|---|---|---|---|
| 2 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3 | 7 | |
| 3 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 1122 | 1564 | |
| 4 | or/1-3 | 1230 | 1727 | |
| 5 | exp clinical trial/ | 401 | 521 | Concept clinical trials |
| 6 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 138560 | 174420 | |
| 7 | meta-analysis as topic/ | 1 | 0 | Concept meta-analysis |
| 8 | meta-analysis/ | 34 | 99 | |
| 9 | (meta-analy* or metaanaly*).ti,ab,kf. | 27362 | 37834 | |
| 10 | or/5-9 | 156727 | 199473 | Concept clinical trials OR meta-analysis |
| 11 | 4 and 10 | 146 | 191 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| 12 | limit 11 to (english or french or german or spanish) | 145 | 191 | Restriction to English, French, German, Spanish:<br>Final result for non-indexed Medline citations |
| | | | 128 | Update search:<br>limit 12 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters
adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**Cochrane Library** (Wiley):
- Cochrane Database of Systematic Reviews
- Cochrane Protocols
- Cochrane Central Register of Controlled Trials

| ID | Search | Annotations |
|---|---|---|
| #1 | ((data near share*) or (data near sharing*)):ti,ab,kw | Concept data sharing: Textword search in title, abstract, keywords |
| #2 | (data next share*) or (data next sharing*) | Concept data sharing: Textword search in fulltext |

| | | |
|---|---|---|
| #3 | #1 or #2 | Concept data sharing. 1. Interim result |
| #4 | ((patient* or participant*) near individual*):ti,ab,kw | Concept Individual patient data sharing. 2. Interim result |
| #5 | data:ti,ab,kw | |
| #6 | (share* or sharing*):ti,ab,kw | |
| #7 | #4 and #5 and #6 | |
| #8 | (IPD near (share* or sharing*)):ti,ab,kw | Concept IPD sharing: 3. Interim result |
| #9 | #3 or #7 or #8 in Cochrane Reviews, Cochrane Protocols, Trials | OR-combination of interim results. Limit to Cochrane Reviews, Protocols, Trials: final result |

| Results | Initial search<br>Search date 2018-10-29 | Update search<br>Publication Year 2018-2020<br>Search date 2020-09-11 |
|---|---|---|
| Cochrane Reviews | 19<br>Issue 10 of 12, October 2018 | 12<br>Issue 9 of 12, September 2020 |
| Cochrane Protocols | 1<br>Issue 10 of 12, October 2018 | 0<br>Issue 9 of 12, September 2020 |
| Trials | 416<br>Issue 9 of 12, September 2018 | 268<br>Issue 9 of 12, September 2020 |

ti,ab,kw = title,abstract keywords
near = terms in any order (default: within 6 words)
next = phrase searching: terms next to each other in the given order
* = truncation

Via Web of Science (Clarivate Analytics):
Databases:
**- Science Citation Index Expanded** (SCI-EXPANDED): 1945-present
**- Social Sciences Citation Index** (SSCI): 1956-present

| **Set** | **Query** | **Results Initial search**<br>Search date: 2018-10-29<br>**Time span: all years**<br>Data last updated: 2018-10-26 | **Results update search**<br>Search date: 2020-09-11<br>**Timespan=2018-2020**<br>Data last updated: 2020-09-10 | **Annotations** |
|---|---|---|---|---|
| # 1 | ts=(("data" near/3 share*) or ("data" near/3 sharing*)) | 12,398 | 5,227 | Concept data sharing |

| | | | | |
|---|---|---|---|---|
| # 2 | ts=((patient* or participant*) near/3 individual*) | 73,929 | 17,281 | Concept Individual patient data sharing |
| # 3 | ts="data" | 5,101,598 | 1,087,419 | |
| # 4 | ts=(share* or sharing*) | 456,989 | 114,300 | |
| # 5 | #4 AND #3 AND #2 | 714 | 336 | |
| # 6 | ts=("IPD" near/6 (share* or sharing*)) | 23 | 24 | Concept IPD sharing |
| # 7 | #6 OR #5 OR #1 | 12,970 | 5,478 | OR-combination of concepts |
| # 8 | ts=(randomi?ed or "randomly" or randomi?ation) | 1,044,391 | 201,056 | Concept clinical trials |
| # 9 | ts=((random* or "clinical") near/3 trial*) | 773,198 | 154,885 | |
| # 10 | ts=("meta analy*" or metaanaly*) | 313,038 | 105,276 | Concept meta-analysis |
| # 11 | #10 or #9 OR #8 | 1,478,458 | 323,705 | Concept clinical trials OR meta-analysis |
| # 12 | #11 AND #7 | 1,022 | 453 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| # 13 | #11 AND #7<br>**Refined by: DOCUMENT TYPES:** ( ARTICLE OR REVIEW ) | 862 | 419 | Restriction to Article or Review: final result |

ts = topic: Title, Abstract, Author Keywords, Keywords Plus®
near/n = terms in any order within n words
* = truncation
? = wildcard for exact 1 character

# Supplementary material 3: Study characteristics

| Author | Year | Country | Type of research | Detail if type of research=other | Type of shared material |
|---|---|---|---|---|---|
| **Tudur-Smith C et al.** | 2014 | UK | Survey | | IPD |
| **Murugiah K et al.** | 2016 | US | Survey | | IPD |
| **Krleža-Jerić K et al.** | 2009 | Canada | Survey | | IPD |
| **Jones CW et al.** | 2016 | US | Survey | | IPD |
| **Mayo-Wilson E et al.** | 2015 | US | Other | Case study | IPD |
| **Reidpath DD et al.** | 2001 | Australia | Experim. | | IPD |
| **Chalmers I et al.** | 2013 | UK | Other | Case study | IPD |
| **Bergeris A et al.** | 2018 | US | Survey | | IPD |
| **Tudur Smith C et al.** | 2017 | UK | Other | Case study | Broader |
| **Vaduganathan M et al.** | 2018 | US | Metrics | | IPD |
| **Merson L et al.** | 2015 | Vietnam | Qualitative | | IPD |
| **Rowhani-Farid A et al.** | 2016 | Australia | Survey | | IPD |
| **Rathi V et al.** | 2012 | US | Survey | | IPD |
| **Ali J et al.** | 2015 | US | Survey | | IPD |
| **Hopkins C et al.** | 2016 | UK | Survey | | IPD |
| **Sydes M et al.** | 2015 | UK | Metrics | Case study | IPD |
| **Polanin J et al.** | 2019 | US | Experim. | | IPD |
| **Villain B et al.** | 2015 | France | Survey | | IPD |
| **Asare A et al.** | 2016 | US | Metrics | | IPD |
| **Strom B et al.** | 2016 | US | Other | Metrics + survey | IPD |
| **Mello M et al.** | 2005 | US | Survey | | IPD |
| **Rathi V et al.** | 2014 | US | Survey | | IPD |
| **Huser V et al.** | 2018 | US | Metrics | | IPD |
| **Chapman S et al.** | 2014 | UK | Survey | | IPD |
| **Griswold M et al.** | 2013 | US | Survey | | Broader |

| | | | | | |
|---|---|---|---|---|---|
| **Cheah PY et al.** | 2015 | Thailand | Qualitative | | IPD |
| **Hee SW et al.** | 2016 | UK | Other | Case study | IPD |
| **Geifman N et al.** | 2015 | US | Metrics | | IPD |
| **Strom B et al.** | 2014 | US | Metrics | | IPD |
| **Ross J et al.** | 2016 | US | Survey | | IPD |
| **Boutron I et al.** | 2016 | France | Survey | | Broader |
| **Vidal-Infer A et al.** | 2018 | Spain | Survey | | Broader |
| **Krumholz H et al.** | 2016 | US | Metrics | | IPD |
| **Tannenbaum S et al.** | 2018 | US | Survey | | IPD |
| **Ross J et al.** | 2017 | US | Metrics | | IPD |
| **Mello M et al.** | 2018 | US | Survey | | IPD |
| **Chickramane A et al.** | 2017 | India | Survey | | IPD |
| **Howe N et al.** | 2018 | UK | Qualitative | | IPD |
| **Naudet F et al.** | 2018 | US | Survey | | IPD |
| **Yuanyuan J et al.** | 2017 | China | Survey | | IPD |
| **Spence O et al.** | 2018 | US | Survey | | IPD |
| **Polanin J et al.** | 2018 | US | Survey | | IPD |
| **Zhu C et al.** | 2017 | US | Metrics | | IPD |
| **So D et al.** | 2017 | Canada | Survey | | IPD |
| **Savage C et al.** | 2009 | US | Survey | | IPD |
| **Kawahara T et al.** | 2018 | Japan | Metrics | | IPD |
| **Goldacre B et al.** | 2017 | UK | Survey | | IPD |
| **Pisani E et al.** | 2017 | UK | Other | Metrics + Qualitative research | IPD |
| **Bertagnolli M et al.** | 2017 | US | Metrics | | IPD |
| **Coady S et al.** | 2017 | US | Metrics | | IPD |
| **Hopkins A et al.** | 2018 | Australia | Survey | Survey | IPD |
| **Piwowar H et al.** | 2007 | US | Survey | | IPD |
| **Laine C et al.** | 2009 | US | Survey | | Broader |
| **Shmueli-Blumberg D et al.** | 2013 | US | Metrics | | IPD |
| **de Vito N et al.** | 2018 | UK | Survey | | Broader |

| | | | | |
|---|---|---|---|---|
| **Nevitt S et al.** | 2017 | UK | Survey | IPD |
| **Ahmed I et al.** | 2011 | UK | Survey | IPD |
| **Navar A et al.** | 2016 | USA | Metrics | IPD |
| **Ebrahim S et al.** | 2014 | USA | Survey | IPD |
| **Vassar M et al.** | 2020 | USA | Survey | IPD |
| **Cheah PY et al.** | 2018 | Thailand | Qualitative | IPD |
| **Staham EE et al.** | 2020 | USA | Survey | Broader |
| **Nutu D et al.** | 2019 | Romania | Survey | IPD |
| **Aleixandre-Benavent R et al.** | 2019 | Spain | Survey | IPD |
| **Ross JS et al.** | 2018 | USA | Metrics | Broader |
| **Gorman DM et al.** | 2019 | USA | Survey | IPD |
| **Bosserdt M et al.** | 2019 | Germany | Survey | IPD |
| **Whitlock EP et al.** | 2019 | USA | Survey | IPD |
| **Gabelica M et al.** | 2019 | Croatia | Survey | IPD |
| **Gorman DM et al.** | 2020 | USA | Survey | IPD |
| **Kaufmann I et al.** | 2019 | UK | Survey | IPD |
| **Veroniki AA et al.** | 2019 | Greece | Experim. | IPD |
| **Godolphin PJ et al.** | 2019 | UK | Experim. | Broader |
| **Rowhani-Farid A et al.** | 2020 | USA | Experim. | IPD |
| **Siebert M et al.** | 2020 | France | Survey | IPD |
| **Mayer C et al.** | 2019 | USA | Survey | IPD |
| **Gaba JF et al.** | 2020 | France | Survey | Broader |
| **Colombo C et al.** | 2019 | Italy | Survey | IPD |
| **Kochhar S et al.** | 2019 | India | Metrics | IPD |
| **Broes S et al.** | 2020 | Belgium | Qualitative | IPD |
| **Rollando P et al.** | 2020 | France | Survey | Broader |
| **Schmidt H et al.** | 2018 | Germany | Metrics | IPD |
| **Azar M et al.** | 2020 | Canada | Survey | IPD |
| **Almaqrami BS et al.** | 2020 | China | Survey | IPD |

| | | | | |
|---|---|---|---|---|
| **Papageorgiou SN et al.** | 2019 | Switzerland | Survey | iPD |
| **Miller J et al.** | 2019 | USA | Survey | Broader |
| **Lovato L et al.** | 2018 | USA | Metrics | IPD |
| **Kemper JM et al.** | 2020 | Australia | Survey | IPD |
| **Johnson AL et al.** | 2020 | USA | Survey | Broader |
| **Sherry C et al.** | 2019 | USA | Survey | Broader |
| **Pellen C et al.** | 2020 | France | Survey | Broader |
| **Danchev V et al.** | 2020 | USA | Survey | IPD |
| **Li R et al.** | 2020 | USA | Metrics | IPD |

Broader: the definition is not solely restricted to IPD and can cover other type of additional material (e.g. protocol, code, etc).

# Supplementary material 4: Published studies about YODA and CSDR

| Reference | Repository/ platform | No. of trials included in repository/ platform | No. of requests | No. of access to data | No. of publications | Date of assessment |
|---|---|---|---|---|---|---|
| Ross, 2017 | YODA | 189 | 73 | 50 | 2 | 6/2017 |
| Vadugan athan, 2018 | YODA | 537 | | 30 | 3 | 5/2017 |
| Strom, 2016 | CSDR | 237 | 177 | 144 | 1* | 11/2015 |

*based upon an investigator survey with 24 responses

# Supplementary material 5: Published outputs from YODA (up to 1st July 2019) and CSDR (up to 31 August 2019)

| Published outputs | Title | Platform used | Identification of the proposal | Type of study | Request date |
|---|---|---|---|---|---|
| **Allott EH et al. 2017** | Statin Use, Serum Lipids, and Prostate Inflammation in Men with a Negative Prostate Biopsy: Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Smoking Is Associated with Acute and Chronic Prostatic Inflammation: Results from the REDUCE Study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Branche BL et al. 2017** | Sleep Problems are Associated with Development and Progression of Lower Urinary Tract Symptoms: Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Vidal AC et al. 2016** | Racial differences in prostate inflammation: results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2016** | Does Prostate Size Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2017** | Does Peak Urine Flow Rate Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Chronic baseline prostate inflammation is associated with lower tumor volume in men with prostate cancer on repeat biopsy: Results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Kent DM et al. 2016** | Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. | CSDR | 647 | Methodological | 29/10/2013 |
| **Baay M et al. 2017** | Background rates of disease in Latin American children from a rotavirus vaccine study. | CSDR | 651 | Secondary analysis | 11/03/2014 |
| **Le Noury J et al. 2015** | Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. | CSDR | 669 | Re-analysis | 27/01/2014 |
| **Nevitt SJ et al. 2017** | Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. | CSDR | 674 | Methodological | 15/05/2014 |
| **Nevitt SJ et al. 2017** | Antiepileptic drug monotherapy for epilepsy: a network meta-analysis of individual participant data. | CSDR | 674 | Meta-analysis | 15/05/2014 |

| | | | | | |
|---|---|---|---|---|---|
| **Forbess LJ et al. 2017** | Failure of a systemic lupus erythematosus response index developed from clinical trial data: lessons examined and learned. | CSDR | 911 | Secondary analysis | 25/07/2014 |
| **Dennis JM et al. 2018** | Evaluating associations between the benefits and risks of drug therapy in type 2 diabetes: a joint modeling approach. | CSDR | 930 | Secondary analysis | missing |
| **Dennis JM et al. 2018** | Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data. | CSDR | 930 | Secondary analysis | missing |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 945 | Secondary analysis | 23/04/2014 |
| **Mistry HB et al. 2017** | Model based analysis of the heterogeneity in the tumour size dynamics differentiates vemurafenib, dabrafenib and trametinib in metastatic melanoma. | CSDR | 946 | Secondary analysis | 28/05/2014 |
| **Muff S et al. 2018** | Bias away from the null due to miscounted outcomes? A case study on the TORCH trial. | CSDR | 977 | Re-analysis | 12/05/2014 |
| **Fragoso CAV et al. 2018** | Spirometric Criteria for Chronic Obstructive Pulmonary Disease in Clinical Trials of Pharmacotherapy. | CSDR | 993 | Secondary analysis | 28/02/2017 |
| **Devilliers H et al. 2016** | Minimal Clinically Important Differences for Generic Patient Reported Outcomes Tools in SLE | CSDR | 998 | Secondary analysis | missing |
| **Li-Kim-Moy J et al. 2018** | Impact of Fever and Antipyretic Use on Influenza Vaccine Immune Reponses in Children. | CSDR | 1000 | Secondary analysis | 08/09/2014 |
| **Blanco JR et al. 2017** | Impact of dolutegravir and efavirenz on immune recovery markers: results from a randomized clinical trial. | CSDR | 1028 | Secondary analysis | 23/09/2014 |
| **Borges NA et al. 2016** | Nonnucleoside Reverse-transcriptase Inhibitor- vs Ritonavir-boosted Protease Inhibitor-based Regimens for Initial Treatment of HIV Infection: A Systematic Review and Metaanalysis of Randomized Trials. | CSDR | 1058 | Meta-analysis | 18/08/2014 |
| **Dodd S et al. 2018** | Incidence and characteristics of the nocebo response from meta-analyses of the placebo arms of clinical trials of olanzapine for bipolar disorder. | CSDR | 1078 | Meta-analysis | 09/10/2014 |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 1079 | Secondary analysis | 12/10/2014 |
| **Emamikia S et al. 2017** | Relationship between glucocorticoid dose and adverse events in systemic lupus erythematosus: data from a randomized clinical trial. | CSDR | 1084 | Secondary analysis | 20/02/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Gruber JF et al. 2018** | Timing and predictors of severe rotavirus gastroenteritis among unvaccinated infants in low- and middle-income countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Gruber JF et al. 2018** | Timing of Rotavirus Vaccine Doses and Severe Rotavirus Gastroenteritis Among Vaccinated Infants in Low- and Middle-income Countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Schwartz LM et al. 2016** | Rotavirus vaccine effectiveness in low-income settings: An evaluation of the test-negative design. | CSDR | 1090 | Secondary analysis | 15/04/2015 |
| **Hilkens NA et al. 2016** | Blood pressure levels and the risk of intracerebral hemorrhage after ischemic stroke. | CSDR | 1100 | Secondary analysis | 13/01/2015 |
| **Hieronymus F et al. 2017** | Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. | CSDR | 1103 | Meta-analysis | missing |
| **Waljee AK et al. 2018** | Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. | CSDR | 1136 | Secondary analysis | 13/08/2015 |
| **Hadjichrysanthou C et al. 2016** | Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. | CSDR | 1137 | Secondary analysis | 16/04/2015 |
| **Voysey M et al. 2017** | The Influence of Maternally Derived Antibody and Infant Age at Vaccination on Infant Vaccine Responses : An Individual Participant Meta-analysis. | CSDR | 1141 | Meta-analysis | 22/07/2015 |
| **Radua J et al. 2017** | Meta-Analysis of the Risk of Subsequent Mood Episodes in Bipolar Disorder. | CSDR | 1148 | Meta-analysis | 30/01/2015 |
| **de Vries YA et al. 2018** | Initial severity and antidepressant efficacy for anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: An individual patient data meta-analysis. | CSDR | 1173 | Meta-analysis | 30/06/2015 |
| **Zafack JG et al. 2019** | Adverse events following immunisation with four-component meningococcal serogroup B vaccine (4CMenB): interaction with co-administration of routine infant vaccines and risk of recurrence in European randomised controlled trials. | CSDR | 1224 | Secondary analysis | missing |
| **Sturm A et al. 2017** | Evaluating the Hierarchical Structure of ADHD Symptoms and Invariance Across Age and Gender. | CSDR | 1292 | Methodological | 29/07/2015 |
| **Oon S et al. 2019** | Lupus Low Disease Activity State (LLDAS) discriminates responders in the BLISS-52 and BLISS-76 phase III trials of belimumab in systemic lupus erythematosus. | CSDR | 1320 | Secondary analysis | missing |
| **Craig K et al. 2017** | More of what works: Detection of informative sites during the conduct of clinical trials using machine learning | CSDR | 1323 | Methodological | 21/10/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Bauza C et al. 2018** | Determining the Joint Effect of Obesity and Diabetes on All-Cause Mortality and Cardiovascular-Related Mortality following an Ischemic Stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Bauza C et al. 2018** | Determining the joint effect of obesity and diabetes on functional disability at 3-months and on all-cause mortality at 1-year following an ischemic stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Tajgardoon M et al. 2018** | A Novel Representation of Vaccine Efficacy Trial Datasets for Use in Computer Simulation of Vaccination Policy. | CSDR | 1374 | Secondary analysis | 25/05/2016 |
| **Berenguer J et al. 2019** | Mathematical modeling of HIV-1 transmission risk from condomless anal intercourse in HIV-infected MSM by the type of initial ART. | CSDR | 1403 | Secondary analysis | missing |
| **Hilkens NA et al. 2017** | Predicting Major Bleeding in Ischemic Stroke Patients With Atrial Fibrillation. | CSDR | 1455 | Secondary analysis | 03/06/2016 |
| **Kerr SJ et al. 2017** | The FDA snapshot algorithm may overestimate the efficacy of initial art | CSDR | 1456 | Methodological | missing |
| **Samara MT et al. 2017** | Initial symptom severity of bipolar I disorder and the efficacy of olanzapine: a meta-analysis of individual participant data from five placebo-controlled studies. | CSDR | 1457 | Meta-analysis | 08/06/2016 |
| **Hopkins AM et al. 2018** | Risk Factors for Severe Diarrhea with an Afatinib Treatment of Non-Small Cell Lung Cancer: A Pooled Analysis of Clinical Trials. | CSDR | 1475 | Meta-analysis | missing |
| **Peters EM et al. 2018** | Melancholic Symptoms in Bipolar II Depression and Responsiveness to Lamotrigine in an Exploratory Pilot Study. | CSDR | 1569 | Secondary analysis | 01/11/2016 |
| **de Vries YA et al. 2018** | Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: individual patient data meta-analysis. | CSDR | 1575 | Meta-analysis | 11/10/2016 |
| **Gemeinsamer Bundesausschuss, 2019** | Nutzenbewertungsverfahren zum Wirkstoff Sitagliptin | CSDR | 1593 | Secondary analysis | 04/11/2016 |
| **Hopkins AM et al. 2019** | Effect of early adverse events on response and survival outcomes of advanced melanoma patients treated with vemurafenib or vemurafenib plus cobimetinib: A pooled analysis of clinical trial data. | CSDR | 1599 | Meta-analysis | missing |
| **Carbon M et al. 2018** | Tardive dyskinesia risk with first- and second-generation antipsychotics in comparative randomized controlled trials: a meta-analysis. | CSDR | 1624 | Meta-analysis | missing |

| | | | | | |
|---|---|---|---|---|---|
| **Schwarzman LS et al. 2018** | The Association of Previous Prostate Biopsy Related Complications and the Type of Complication with Patient Compliance with Rebiopsy Scheme. | CSDR | 1626 | Secondary analysis | 16/12/2016 |
| **Voysey M et al. 2017** | Prevalence and decay of maternal pneumococcal and meningococcal antibodies: A meta-analysis of type-specific decay rates. | CSDR | 1629 | Meta-analysis | 26/09/2016 |
| **Shapiro W et al. 2018** | Salmeterol Combined with Fluticasone Reduces Exacerbations More Effectively in Chronic Bronchitis Associated with Chronic Obstructive Pulmonary Disease: A Post-hoc Analysis of the TORCH Trial | CSDR | 1640 | Secondary analysis | 28/02/2017 |
| **Parodis I et al. 2018** | Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Established organ damage reduces belimumab efficacy in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Predictors of low disease activity and clinical remission following belimumab treatment in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hernández-Breijo B et al. 2019** | Antimalarial agents diminish while methotrexate, azathioprine and mycophenolic acid increase BAFF levels in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hopkins AM et al. 2019** | Predictors of Long-Term Disease Control and Survival for HER2-Positive Advanced Breast Cancer Patients Treated With Pertuzumab, Trastuzumab, and Docetaxel. | CSDR | 1741 | Secondary analysis | missing |
| **Janciauskiene S et al. 2019** | Serum Levels of Alpha1-antitrypsin and Their Relationship With COPD in the General Spanish Population. | CSDR | 2084 | Secondary analysis | missing |
| **Storgaard H et al. 2016** | Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis. | YODA | 2014-0340 | Meta-analysis | 19/11/2014 |
| **Christian KE et al. 2019** | Gender Differences and Other Factors Associated with Weight Gain Following Initiation of Infliximab: A Post Hoc Analysis of Clinical Trials. | YODA | 2014-0334 | Meta-analysis | 26/11/2014 |
| **Wang R et al. 2018** | Comparative Efficacy of Tumor Necrosis Factor-α Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis. | YODA | 2014-0291 | Meta-analysis | 08/12/2014 |
| **Waljee AK et al. 2017** | External Validation of a Thiopurine Monitoring Algorithm on the SONIC Clinical Trial Dataset. | YODA | 2014-0401 | Secondary analysis | 20/01/2015 |
| **Mospan GA et al. 2017** | 5-Day versus 10-Day Course of Fluoroquinolones in Outpatient Males with a Urinary Tract Infection (UTI). | YODA | 2015-0514 | Secondary analysis | 26/05/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Singh S et al. 2018** | Impact of Obesity on Short- and Intermediate-Term Outcomes in Inflammatory Bowel Diseases: Pooled Analysis of Placebo Arms of Infliximab Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Singh S et al. 2018** | Obesity and Response to Infliximab in Patients with Inflammatory Bowel Diseases: Pooled Analysis of Individual Participant Data from Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Spertus J et al. 2018** | Risk of weight gain for specific antipsychotic drugs: a meta-analysis. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Spertus J et al. 2019** | Bayesian Meta-analysis of Multiple Continuous Treatments with Individual Participant-Level Data: An Application to Antipsychotic Drugs. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Zou X et al. 2018** | The role of PANSS symptoms and adverse events in explaining the effects of paliperidone on social functioning: a causal mediation analysis approach. | YODA | 2016-0716 | Secondary analysis | 24/02/2016 |
| **World Health Organization 2017** | WHO report (appendix) | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Mbuagbaw L et al. 2019** | Outcomes of Bedaquiline Treatment in Patients with Multidrug-Resistant Tuberculosis. | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Corbett M et al. 2017** | Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. | YODA | 2016-0897 | Meta-analysis | 19/05/2016 |
| **Gay HC et al. 2017** | Feasibility, Process, and Outcomes of Cardiovascular Clinical Trial Data Sharing: A Reproduction Analysis of the SMART-AF Trial. | YODA | 2016-0912 | Re-analysis | 07/06/2016 |
| **Schneider-Thoma J et al. 2018** | Second-generation antipsychotic drugs and short-term mortality: a systematic review and meta-analysis of placebo-controlled randomised controlled trials. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Schneider-Thoma J et al. 2019** | Second-generation antipsychotic drugs and short-term somatic serious adverse events: a systematic review and meta-analysis. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Teply BA et al. 2019** | Risk of development of visceral metastases subsequent to abiraterone vs placebo: An analysis of mode of radiographic progression in COU-AA-302. | YODA | 2016-1057 | Secondary analysis | 01/09/2016 |
| **Loubersac T et al. 2019** | Neutrophil-to-lymphocyte Ratio as a Predictive Marker of Response to Abiraterone Acetate: A Retrospective Analysis of the COU302 Study. | YODA | 2016-1103 | Secondary analysis | 23/11/2016 |
| **Martin LJ et al. 2018** | Identification of subgroups of metastatic castrate-resistant prostate cancer (mCRPC) patients treated with abiraterone | YODA | 2016-1122 | Secondary analysis | 23/11/2016 |

| | | | | | |
|---|---|---|---|---|---|
| | plus prednisone at low- vs. high-risk of radiographic progression: An analysis of COU-AA-302. | | | | |
| **Waljee AK et al. 2019** | Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. | YODA | 2016-1176 | Secondary analysis | 01/03/2017 |
| **Kubo K et al. 2018** | Placebo effects in adult and adolescent patients with schizophrenia: combined analysis of nine RCTs. | YODA | 2017-1676 | Meta-analysis | 24/05/2017 |
| **Kumagai F et al. 2018** | Early Placebo Improvement Is a Marker for Subsequent Placebo Response in Long-Acting Injectable Antipsychotic Trials for Schizophrenia: Combined Analysis of 4 RCTs. | YODA | 2017-1701 | Meta-analysis | 01/06/2017 |
| **Yiu ZZN et al. 2019** | A standardization approach to compare treatment safety and effectiveness outcomes between clinical trials and real-world populations in psoriasis. | YODA | 2017-1706 | Methodological | 08/08/2017 |
| **Narula N et al. 2018** | Patient-Reported Outcomes and Endoscopic Appearance of Ulcerative Colitis: A Systematic Review and Meta-analysis. | YODA | 2017-2031 | Meta-analysis | 30/08/2017 |
| **Singh S et al. 2018** | No Benefit of Concomitant 5-Aminosalicylates in Patients With Ulcerative Colitis Escalated to Biologic Therapy: Pooled Analysis of Individual Participant Data From Clinical Trials. | YODA | 2017-2306 | Meta-analysis | 25/09/2017 |
| **Singh S et al. 2019** | Efficacy and Speed of Induction of Remission of Infliximab vs Golimumab for Patients With Ulcerative Colitis, Based on Data From Clinical Trials. | YODA | 2018-3121 | Meta-analysis | 21/05/2018 |

# Status, use and impact of sharing Individual Participant data from clinical trials: a scoping review

C. Ohmann et al.

**Scoping Reviews (PRISMA.ScR) Checklist:**

| Section | Item | Covered | Page no. in manuscript |
|---|---|---|---|
| Title | Title | yes | 1 |
| Abstract | Structured summary | yes | 2 |
| Introduction | Rationale | yes | 3 |
| | Objectives | yes | 4 |
| Methods | Protocol and registration | yes | 5 |
| | Eligibility criteria | yes | 5 |
| | Information sources | yes | 6 |
| | Search | yes | 7 |
| | Selection of sources of evidence | yes | 7 |
| | Data charting process | yes | 7 |
| | Data items | yes | 7 |
| | Critical appraisal of individual sources of evidence | yes | 7 |
| | Synthesis of results | yes | 8 |
| Results | Selection of sources of evidence | yes | 8 |
| | Characteristics of sources of evidence | yes | 8 |
| | Critical appraisal within sources of evidence | yes | 9 |
| | Results of individual sources of evidence | yes | 9 |
| | Synthesis of results | yes | 9-20 |
| Discussion | Summary of evidence | yes | 20 |
| | Limitations | yes | 22 |
| | Conclusions | yes | 22 |
| Funding | Funding | yes | 3 |

BMJ Open

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-049228.R2 |
| Article Type: | Original research |
| Date Submitted by the Author: | 17-Jul-2021 |
| Complete List of Authors: | Ohmann, Christian; European Clinical Research Infrastructure Network<br>Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre<br>Siebert, Maximilian; University Rennes, CHU Rennes, CIC 1414 (Centre d'Investigation Clinique de Rennes)<br>Motschall, Edith ; University of Freiburg Faculty of Medicine, Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg,<br>Naudet, Florian; University Rennes, CHU Rennes, INSERM CIC 1414 (Centre d'Investigation Clinique de Rennes) |
| Primary Subject Heading: | Health informatics |
| Secondary Subject Heading: | Health informatics |
| Keywords: | Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Information management < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Status, use and impact of sharing Individual Participant Data from clinical trials: a scoping review

**(Date: 16.07.2021)**

**Authors**

C. Ohmann, ORCID: 0000-0002-5919-1003, christianohmann@outlook.de
D. Moher, ORCID: 0000-0003-2434-4206, dmoher@ohri.ca
M. Siebert, ORCID: 0000-0003-4385-5773, maximiliansiebert91@gmail.com
E. Motschall, ORCID: 0000-0001-7795-242X, motschall@imbi.uni-freiburg.de
F. Naudet, ORCID: 0000-0003-3760-3801, floriannaudet@gmail.com

**Corresponding author:**

Prof. Dr. Christian Ohmann
European Clinical Research Infrastructure Network (ECRIN)
Kaiserswerther Str. 70
40477 Düsseldorf
Germany
Tel.: +49 211 5142490
Mobile: + 49 16097945385
Email: christianohmann@outlook.de

**Co-authors :**

David Moher
Ottawa Hospital Research Institute
Ottawa Methods Centre
Ottawa, On
Canada

Maximilian Siebert, Florian Naudet
Université de Rennes 1
CHU Rennes
INSERM, CIC 1414 (Centre d'Investigation Clinique de Rennes)
Rennes, Bretagne
France

Edith Motschall
Institute of Medical Biometry and Statistics
Faculty of Medicine and Medical Center
University of Freiburg
Germany

**Key words:**

clinical trial, individual participant data, data sharing, scoping review, impact

**Word count:**

13194

**ABSTRACT**

**Objectives**
To explore the impact of data-sharing initiatives on the intent to share data, on actual data-sharing, on the use of shared data and on research output and impact of shared data.
**Eligibility criteria**
All studies investigating data-sharing practices for individual participant data (IPD) from clinical trials.
**Sources of evidence**
We searched the Medline database, the Cochrane Library, the Science Citation Index Expanded and the Social Sciences Citation Index via Web of Science, and preprints and proceedings of the International Congress on Peer Review and Scientific Publication. In addition, we inspected major clinical trial data-sharing platforms, contacted major journals/publishers, editorial groups and some funders.
**Charting methods**
Two reviewers independently extracted information on methods and results from resources identified using a standardised questionnaire. A map of the extracted data was constructed and accompanied by a narrative summary for each outcome domain.
**Results**
93 studies identified in the literature search (published between 2001-2020, median: 2018) and 5 from additional information sources were included in the scoping review. Most studies were descriptive and focused on early phases of the data-sharing process. While the willingness to share IPD from clinical trials is extremely high, actual data-sharing rates are suboptimal. A survey of journal data suggests poor to moderate enforcement of the policies by publishers. Metrics provided by platforms suggest that a large majority of data remains unrequested. When requested, the purpose of the re-use is more often secondary analyses and meta-analyses, rarely re-analyses. Finally, studies focused on the real impact of data-sharing were rare and used surrogates such as citation metrics.
**Conclusions**
There is currently a gap in the evidence base for the impact of IPD sharing, which entails uncertainties in the implementation of current data-sharing policies. High level evidence is needed to assess whether the value of medical research increases with data-sharing practices.

**Strengths and limitations of this study**

- Exhaustive review of both the literature and the main initiatives in data-sharing

- Analysis of the full data-sharing process covering intention to share, actual sharing, use of shared data, research output and impact

- Retrieval and synthesis of information proved to be difficult because of a very siloed landscape where each initiative/platform operates with its own metrics

- Data-sharing is a moving target in a rapidly changing environment with more and more new initiatives.

- Only a limited research output from data-sharing is available so far

- The time from submitting a data sharing request to receiving the requested data was not systematically investigated in the review

**Funding**

No specific funding for this review.
FN's work on data-sharing is supported by a grant from the French National Research Agency – ANR (Reproducibility in Therapeutic Research / ReITheR: Agence Nationale de la Recherche, ANR-17-CE36-0010-01).
CO's work on data-sharing is supported by funding from the European Union's Horizon 2020 Research and Innovation Programme (CORBEL, under grant agreement n° 654248).
DM is supported by an Ottawa University Research Chair (grant number: N/A).

**Competing interests**

None of the authors have any competing interests.

**Author's contribution**

CO, DM and FN developed the study protocol. The search strategy was developed and implemented by EM. The selection of sources of evidence and the assessment was performed by CO and FN. Contact with initiatives/platforms/journals/publishers was performed by MS. In case of disagreements, these were resolved by consensus and, when necessary, in consultation with DM. The first draft of the manuscript was written by CO and FN. All authors have revised and approved the final manuscript.

**Data sharing statement**

All data relevant to the study is included in the article or uploaded as supplementary Information. There is a link to the code and the data on the OSF (https://osf.io/h6cj4/).

**Acknowledgements**

We thank Angela Swaine Verdier for revising the English in the manuscript.

## INTRODUCTION

### Rationale

Data sharing is increasingly recognized as a key requirement in clinical research.[1] In any discussion about clinical trial data-sharing the emphasis is naturally on the data sets themselves, but data-sharing is much broader. Besides the individual participant data (IPD) sets, other clinical trial data sources should be made available for sharing (e.g., protocols, clinical study reports, statistical analysis plans, blank consent forms) to enable a full understanding of any data set. In this scoping review, there is a focus on the sharing of individual participant data from clinical trials.

Within clinical research, data-sharing can enhance reproducibility and the generation of new knowledge, but it also has an ethical and economic dimension.[2] Scientifically, sharing makes it possible to compare or combine the data from different studies, and to more easily aggregate it for meta-analysis. It enables conclusions to be re-examined and verified or, occasionally, corrected, and it can enable new hypotheses to be tested. Sharing can therefore increase data validity, but it also draws more value from the original research investment, as well as helping to avoid unnecessary repetition of studies. Agencies and funders are referring more and more to the economic advantages of data reuse. Ethically, data-sharing provides a better way to honour the generosity of clinical trial participants, because it increases the utility of the data they provide. Despite the high potential for sharing clinical trial data, the launch and implementation of several data-sharing initiatives and platforms, and outstanding examples related to the value of data-sharing,[3] to date data-sharing is not the norm in clinical research, unlike many other scientific disciplines.[4] One major hurdle is that clinical trial data concerns individuals and their health status, and as such requires specific measures to protect privacy.

To support sharing of IPD in clinical trials, several organisations have developed generic principles, guidance and practical recommendations for implementation. In 2016, the International Committee of Medical Journal Editors (ICMJE), a small group of medical journal editors, published an editorial stating that "it is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk". [5] The ICMJE considers that there is an implicit social contract imposing an ethical obligation for the results to lead to the greatest possible benefit to society. The ICMJE proposed to require that de-identified IPD is made publicly available no later than 6 months after publication of the main trial results. This time lapse would be useless for public health emergencies like COVID-19. However, the ICMJE proposal triggered debate, and a large number of trialists were reluctant to adopt this new norm[6] on account of the feasibility of the proposed requirements, the resources required, the real or perceived risks to trial participants, and the need to protect the interests of patients and researchers.[7]

Despite the cultural shift towards sharing clinical trial data and the major commitment of scientific organisations, funders and initiatives, overall there is still a lack of effective policies in the biomedical literature to ensure that underlying data is maximally available and reusable. The only requirement appears to be a data management plan or a data-sharing plan. A few journals require data-sharing and, for those who do require data-sharing, guidelines are heterogeneous and somewhat ambiguous.[8] Nevertheless, some innovative and progressive funders (e.g. Wellcome Trust, Bill & Melinda Gates Foundation), and publishers/journals (e.g. Public Library of Science (PLOS) [in 2014], The British Medical Journal (BMJ)) [2009-2015], have adopted strong data-sharing policies. As part of a wider cultural shift towards more open science, there have been various attempts to explore how clinical researchers can best plan for data-sharing and prepare their 'raw' IPD so that it becomes available to others[9] – albeit often under controlled access conditions rather than simply being publicly available on-line[10] - and can structure that data to make it FAIR (findable, accessible, interoperable and reusable).[11] Meanwhile several data-sharing platforms and repositories are available and in use to provide practical support for the data-sharing process in clinical research (e.g. Yale University Open Data Access (YODA) launched [in 2011], ClinicalStudyDataRequest.com (CSDR) [launched in 2013], Vivli [launched in 2018]. A considerable number of individual studies have been performed to access and explore the sharing of data from clinical trials under different circumstances and within different frameworks. What is strongly needed is a scoping review providing an overview of the status of implementation of data-sharing as a whole and the implications originating from the available evidence.

### Objectives

In this scoping review we explored the impact of data-sharing initiatives on the willingness to share data, the status of data-sharing, the use of shared data and the impact of research outputs from shared data.

## METHODS

### Protocol and registration

The study protocol was registered on the Open Science Framework on September the 12th 2018 (registration number: osf.io/pb8cj). The protocol followed the methodology manual published by the Joanna Briggs Institute for scoping reviews.[12] Methods and results are reported using the PRISMA (Preferred Reporting Items for systematic Reviews and Meta Analyses) extension for scoping reviews (PRISMA-ScR).[13]

### Eligibility criteria

The following eligibility criteria for studies were used:

All study designs were eligible, including case studies, surveys, metrics and experimental studies, using qualitative or quantitative methods. Only published or unpublished reports (e.g. pre-prints, congress presentations, non-indexed information such as websites) in English, German, French or Spanish were considered.

We included all studies and reports 1/ providing information on current IPD data-sharing practices for clinical trials and 2/ reporting on one or more of five outcome domains defined according to the data-sharing process presented in **Box 1**.

**1. Intention to share data**

There is an intention to share data, expressed by a stakeholder (e.g., sponsor/PI, funder). This can be done by a written data-sharing commitment or by a declaration included in the trial registration. This also includes surveys on attitudes towards data-sharing.

**2. Actual data-sharing**

Data is truly made available for data-sharing to secondary users. This is important because there are cases known where the data is offered for sharing but sharing does not take place, as a result of a possible hidden agenda or change in plans.

**3. Use of shared data**

Shared data can be used for various purposes. It can be used as background for research, usually not leading to research outputs. This covers use for education, researcher training and understanding of data. Study types that should lead to new research outputs include 1/ validation/reproducibility of results, 2/ further additional analyses (prognostic models, decision-support, subgroup analyses, etc.) and 3/ IPD meta-analyses.

**4. Research outputs from shared data**

Research outputs are scientific presentations, reports and publications.

**5. Impact of research output from shared data**

Research output from shared data can have an impact on medical research (e.g. development of new hypotheses and methods) and/or medical health (e.g. changes in treatment via guidelines).

**Box 1:** Definitions used for the 5 outcome domains

In the scoping review only data-sharing of IPD from clinical trials was considered. We defined clinical trials following the ClinicalTrials.gov definition: "a clinical study is a research study involving human volunteers (also called participants) that is intended to add to medical knowledge. There are two types of clinical studies: interventional studies (also called clinical trials) and observational studies. Clinical trial is another name for an interventional study."[14] We therefore considered any interventional clinical studies (no matter whether they were randomised), and we did not consider studies on data-sharing concerning observational and non-clinical studies (e.g. on genomics) nor different fields outside medicine (e.g. economics).
We included studies that investigated and reported information on current data-sharing practices performed without restrictions in terms of promotional initiatives, type of repository or platform (see **Box 2** for definitions) and that promoted data-sharing practices (e.g. at editorial level, at funder level, at research level etc.). We considered many different types of studies (e.g. experimental studies, surveys, metrics, quality assurance studies, qualitative research, reviews, reports), as the inclusion criteria were not method-specific but rather content-specific.

**Initiatives**

Major activities of an organization (or a network of several organizations) to actively promote data-sharing in this area (e.g. Pharmaceutical Research and Manufacturers of America (PHRMA)/European Federation of Pharmaceutical Industries and Associations (EFPIA), Nordic Trial Alliance, Institute of Medicine (IOM), ICMJE, Research Data Alliance (RDA)).

**Repository**

Large database infrastructures set up to manage, share, access and archive researchers' datasets from clinical trials. Repositories can be specialised and dedicated to specific disciplines (e.g. FreeBird, Biological Specimen and Data Repository Information Coordination Center (BioLINCC) or more general (e.g. FigShare, Dryad).

**Platform**

A computer environment where researchers can find datasets from clinical trials across different repositories, and where additional functionalities (e.g. protected analysis environment) are provided (e.g. CSDR, YODA, Project Data Sphere, Github).

**Box 2:** Definitions used for initiatives, repository and platform

**Information sources**

The identification of studies was performed in two complementary stages:

a) A systematic literature search in bibliographic databases (MEDLINE databases, Cochrane Library, Science Citation Index Expanded and Social Science Citation Index). In addition, preprint servers and proceedings were searched
b) Inspection of and if required contacts with known information sources (e.g. webpages, documents and reports from platforms, funder, publisher) to explore whether they had an evaluation component and provided detailed research output from shared data (see supplementary material 1).

Between 25/01/2019 and 12/06/2019 (with an update on 02/11/2020), one researcher (MS) inspected (and when necessary contacted) major clinical trial data-sharing platforms to explore whether they had an evaluation component and provided details of research output from shared data (see **Supplementary Material 1**). Similarly, in the same time period, the researcher contacted major journals and/or publishers and/or editorial groups (The BMJ, PLOS, The Annals of Internal Medicine, BioMedCentral (Springer/Nature), Faculty of 1000 Research (F1000Research)). These journals/publishers were targeted because they had either an early or a robust data-sharing policy (NEJM, Lancet and JAMA had no data-sharing policy before the 2018 ICMJE policy). Some funders (see **Supplementary Material 1**) were also contacted, and preprints repositories were explored (bioRxiv, PeerJ, Preprints.org, PsyArXiv and MedRxiv. For the sake of completeness, ASAPbio (Accelerating Science and Publication in biology) and the Center for Open Science were also contacted for the

same information, as well as three International Congress on Peer Review and Scientific Publication conference abstracts. In addition, when relevant references were found in various papers these references were included (snowballing searches).

**Search**

On 29/10/2018 (update on 12/09/2020), one researcher (EM) searched the Medline databases for indexed and non-indexed citations via Ovid from Wolters Kluwer, the Cochrane Library via Wiley, Science Citation Index Expanded and Social Sciences Citation Index via Web of Science from Clarivate Analytics for articles meeting our inclusion criteria.

The detailed search terms for the MEDLINE databases, the Cochrane Library and the Web of Science databases can be found in **Supplementary Material 2**. The main search strategy developed by CO, DM und FN was peer-reviewed independently (by a senior medical documentalist, EM who joined the team subsequently) using evidence-based guidelines for Peer Review of Electronic Search Strategies (PRESS).[15] Discrepancies were resolved between the authors, and EM performed the search. All references were managed and de-duplicated using a reference manager system (Endnote).

On 23/01/2019 (update on 02/11/2020), two researchers (MS and FN) independently searched for relevant pre-prints on OSF PREPRINTS using the search function to find all papers relevant to medicine with the following keyword (trial* OR random*). On 29/01/2019, the two researchers independently searched the proceedings of the three latest International Congress on Peer Review and Scientific Publication reports for relevant abstracts (2009, 2013 and 2017).

**Selection of sources of evidence**

The selection of sources of evidence was performed by two independent reviewers (CO and FN). Contact with initiatives/platforms/journals/publishers was made by a single reviewer (MS). In case of disagreements, these were resolved by consensus between CO and FN and, when necessary, in consultation with a third reviewer (DM).

**Data charting process**

We developed a data collection form and pilot-tested it on 10 randomly selected research papers which were later included in our final study. In case of disagreement, these were resolved by consensus and, when necessary, in consultation with a third reviewer (DM).

**Data items**

For each research paper included according to the selection criteria we extracted: 1/ basic information on the paper (type of study exploring data-sharing practices, authors, year, references, and type of initiative and/or repository and/or platform studied), 2/ information on the material shared (sharing of data, code, programs and material), 3/ whether it reported data about one or more of the five outcomes domains defined box 1, 4/ how these outcome domains were assessed, and 5/ a qualitative description of the main results observed on these outcomes.

For each data-sharing platform, publisher and funder providing detailed research output from shared data, we extracted the following information (authors, date of request, date of publication, type of re-use). We initially planned to describe the scale of re-use in qualitative terms and the observed results of the re-use (i.e. “positive” or “negative” study) but these two characteristics were difficult to extract with very poor inter-rater agreement and we decided not to detail them.

**Critical appraisal of individual sources of evidence**

The studies included were classified according to study type (e.g. survey, metrics, experimental). Potentially relevant characteristics of studies included with regard to their internal-external validity and risk of bias were

not assessed systematically with a specific tool, but explored when one of the two reviewers considered it relevant, and in this case each study was thoroughly discussed between the reviewers.

**Synthesis of results**

No outcome was prioritized since there was no quantitative synthesis for this study. All outcomes were described separately in sections corresponding to the outcome domain and subsections corresponding to similar types of initiative. Our plan for the presentation of results was specified in our protocol and organized into 1/ different sections corresponding to the key concepts detailed in the data-sharing pipeline (intention to share data, actual data-sharing, results of re-use, output from data-sharing, impact of data-sharing) and 2/ different subsections corresponding to the different contexts and actors involved in the data-sharing pipeline (e.g. targeted group for intention to share data or type of use for re-use of shared data)). A summary of the data extracted from the papers included was constructed in tabular form with basic characteristics, and was accompanied by a narrative summary describing all results observed in the light of the review objective and *question/s.* Usually, individual studies were summarized in a short text with descriptive statistics of the main results (numbers, percentages), when appropriate visual representations of the data extracted were provided.

**Patient and public involvement**

There was no patient or public involvement in this scoping review.

**Changes to the initial protocol**

We initially planned to contact leading authors in the field to ask whether they were aware of other unpublished initiatives, but this was not done as it was difficult to identify relevant authors. We found relevant references about data-sharing policies including both clinical trials and observational studies, without making a distinction. These references were included in the scoping review and this point was discussed in the text.

**RESULTS**

**Selection of sources of evidence**

A total of 3024 records were identified, 3,005 records (1991 + 1014 in the update) were retrieved by database search (2141 without duplicates). An additional 8 records were identified by screening the proceedings of the last three International Congress on Peer Review and Scientific Publication conference abstracts and ten records by snowballing searches. One additional relevant record was identified after screening 630 identified pre-prints. We screened all irrelevant records by title and abstract, leaving 409 possibly relevant references which were eligible for full-text screening. Subsequently, 316 references were excluded, leaving 93 reports that met the inclusion criteria (**Figure 1**). We inspected websites and when needed contacted 48 initiatives/platforms/journals (we actually screened 49 but Supporting Open Access for Research Initiative (SOAR) is now integrated into Vivli): 23 data-sharing platforms, 13 funding organisation, 5 journals, 5 pre-print repositories and 2 other initiatives. For 33 of these different sources, there was no evaluation component and for 10 additional contacts we received no answer as to whether they had an evaluation component and/or any data. 4 data-sharing platforms (CSDR, YODA, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Vivli) and 1 funding organisation (Medical Research Council United Kingdom (MRC UK)) provided some additional data (online metrics and or data about its policy) (**Figure 1**) which was extracted in June 2019 and updated in December 2020.

**Characteristics of sources of evidence**

Of the 93 reports, 5 were classified as experimental studies, 58 as surveys, 19 as metrics, 5 as qualitative research and 6 as other (4 case studies, 1 metrics & survey, 1 metrics and qualitative). The median year of publication was 2018 (range [2001-2020]). The vast majority of these studies were from North America (50,

54%), Europe (16, 17 %) and the UK (15, 16%). Eight (9%) were from Asia and 4 (4 %) from Australia. Most (78, 84 %) were focused on IPD data-sharing while the remaining 15 (16 %) adopted a wider definition of the material shared (e.g. by including protocols, codes). Thirty-eight reports (41 %) were focused on data-sharing in publications/journals, 23 (25 %) on data repositories, 8 (9 %) on data-sharing by various institutions, 4 (4 %) on trial registries and 20 (21 %) in various other contexts (see **Supplementary Material 3** which presents study characteristics in detail).

**Collating and summarising the data**

**Figure 2** shows the proportion of the 93 references exploring each outcome domain. In an effort to create a useful synthesis of results, we collated results on each outcome from each publication and organised them into the pre-specified categories. **Figure 3** presents a detailed overview of the different outcome domains and the related outcomes used in the 93 different references included, organised by type of research.

**Critical appraisal of sources of evidence**

In general, there was a high risk of bias, especially due to study design (e.g. surveys with low response rates and absence of experimental design). As stated in the methods, this was not assessed systematically. If available, we have tried to present this information in the narrative part of the review.

**Results for individual sources of evidence: intentions to share data**

***Clinical Trialists***

***Surveys of attitudes***

Four surveys investigating intention to share data by trialists reported high data-sharing rates of around 75% or more (see **Figure 4**). These surveys targeted authors of published trials and in one study reviewers in a Cochrane group (where the majority of respondents had been involved in a randomised controlled trial (RCT)). The studies differed by different estimations of data-sharing rates, different selection criteria and/or survey methods. Response rates were comparable across the surveys (42-58%). Reviewers in the Cochrane IPD meta-analysis group were strongly in favour of a central repository and of providing IPD for central storage (83%)[16]. In the survey by Rathi et al., 74% and 72% respectively thought that sharing de-identified data through data repositories should be required and that investigators should be required to share de-identified data in response to individual requests. However, only 18% indicated that they were required by the trial funder to place the trial data in a repository. In this survey, support for data-sharing did not differ on trialist or trial characteristics. [17] Trialists in Western Europe indicated they had shared or would share data in order to receive academic benefits or recognition more frequently than those from the USA or Canada (58 versus 31%). The most academically productive trialists less frequently indicated they had withheld or would withhold data in order to protect research subjects (24 versus 40% for the least productive), as did those who had received industry funding compared to those who had not (24 versus 43%).[18] The survey by Tannenbaum, 2018 suggested that willingness to share data could depend on the intended re-use of the data (97% of respondents were willing to share data for a meta-analysis versus 73% for a re-analysis).[19] For secondary analyses, the willingness to share was largely influenced by respondents' willingness to conduct a similar analysis. In addition, willingness to share was more marked after 1 year than after 6 months. In the fourth survey on trials published in Chinese medical journals, the overwhelming majority (87%) stated that they endorsed data-sharing.[20]

*Metrics of data-sharing statements in journal articles*

Intentions to share data for trialists were less clear for data-sharing statements in published journal articles (although this section is not specific to clinical trials) (see **Figure 4**). Depending on the journals considered, the rates vary from less than 5 % to around 25%. An analysis of the first year after the Annals of Internal Medicine policies encouraged data-sharing found that data was available without condition for 4%, with conditions for 57%, and unavailable for 38% .[21] Over the first 4 years data was available without condition for 7%, with

conditions for 47%, and unavailable for 46% of research articles.[22] 9% and 22 % of 160 randomly sampled research articles in the BMJ from 2009 to 2015 made data available or indicated the availability of their data sets.[23] Among 60 randomized cardiovascular interventional trials registered on ClinicalTrials.gov, up to 2015 with >5000 enrollment, sponsored by one of the top 20 pharmaceutical companies in terms of 2014 global sales, IPD was available for 15 trials (25%) amounting to 204 452 patients, unavailable for 15 trials (25%) and undetermined for the remaining 50 %, because of either no response or requirements for a full proposal.[24] Reasons for non-availability were: co-sponsor did not agree to make IPD available (4 trials) and trials were not conducted within a specific time (5 trials); for the remaining 6 trials, no specific reason was provided. Of 619 RCTs published between 2014 - 2016 in 7 high-ranked anaesthesiology journals, only 24 (4%) had a data-sharing statement and none provided data in the manuscript or a link to data in a repository.[25] In a survey targeting the authors of these RCTs, 86 (14%) responded and raw data was obtained from 24 participants. The authors conclude that willingness to share data among anaesthesiology RCTs is very low. From 1 July 2018, clinical trials submitted to ICMJE journals are required to contain a data-sharing statement. The reporting of the statement was investigated in a 2-month period before and after this date.[26] The proportion of articles with a data-sharing statement was 23% (32/137) before and 25% (38/150) after 1st July 2018, while the number of journals publishing data-sharing statements increased from 4/11 to 7/11. Few data-sharing statements complied fully with the ICMJE journal criteria, and the majority did not refer to individual participant data. A total of 300 trials published in 2017-2018 and approximately equally distributed across orthodontics and periodontics were selected, assessed, and analysed with respect to transparency and reporting.[27] Open data-sharing (repository or appendix) was found in 5 % of the trials (11/150 orthodontics and 4/150 periodontics trials). Articles on reproducible research practices and transparency in reproductive endocrinology and infertility (REI) were investigated for original articles with a study type mix from REI journals (2013, 2018) and articles published in high-impact general journals between 2013 – 2018.[28] Raw data was available on request or via online database for 1/98 articles in reproductive endocrinology and infertility RCTs (2013), 0/90 in 2018 and 1/34 in high impact journals. In a random sample of 151 empirical studies in 300 otolaryngology research publications, using a PubMed search for records published between 1 January 2014 and 31 December 2018, only 5 provided a data availability statement and 3 (2.0%) indicated that data was available.[29]

*Metrics of data-sharing statements in clinical trial registries*

Intention to share could be even lower when considering data-sharing plans of trials registered at ClinicaTrials.gov. Here the willingness to share data is between 5 and 10%. In one study, 25 551 trial records responded to the Plan to share IPD (72%). Of these, 10.9% of the records indicated "yes" and 25.3% indicated "undecided".[30] Differences were observed by key funder type, with 11% of NIH funders and 0% in the industry answering yes. Importantly, an in-depth review of 154 data-sharing plans suggested a possible misunderstanding of IPD sharing with discrepancies found between data-sharing plans and reports of actual data-sharing. In a survey, the prevalence and quality of IPD-sharing statements among 2,040 clinical trials first posted on ClinicalTrials.gov between 01 January 2018 and 06 June 2018 were investigated.[31] The vast majority of trials included in this study did not indicate an intention to share IPD (n = 1,928; 94.5%). Among the trials that did commit to sharing IPD (n = 112, 5.5%), significant variability existed in the content and structure of the IPD sharing statements with a need for further clarification, enhanced clarification and better outreach. Data from 287 626 clinical trials registered in ClinicalTrials.gov on 20 December 2018 were analysed with respect to sharing of IPD.[32] Overall, 10.8% of trials with a first registration date after December 1 2015 answered "Yes" to plans to share de-identified IPD data. The sharing rate ranged from 0% (biliary tract neoplasms) to 72.2% (meningitis, meningococcal infection) when analysed by disease. For the case of HIV, which was analysed separately, the sharing rate was higher on average (24.5%). In a prediction model, studies that deposit basic summary results on ClinicalTrials.gov, large studies and phase 3 interventional studies are the most likely to declare intention to share IPD data.

*Other data sources*

A 2015 survey focused on PCORnet (The National Patient-Centered Clinical Research Network), found that a possible barrier toward data-sharing intentions related to how data can be used when shared with institutions that have different levels of experience, and to the possibility of some “competition” between institutions on the marketplace of ideas.[33]

*Experimental studies*

Experimental data suggests that estimations of intention to share data could differ depending on the formulation of the request. For instance, a small randomised prospective study conducted in 2001 including 29 corresponding authors of research publications published in the BMJ, explored their preparedness to share the data from their research.[34] The email contact, randomly allocated, was in one of two forms, a general request (asking if the author would "in general" be prepared to release data for re-analysis) and a specific request (a direct request for the data for re-analysis). Researchers receiving specific requests for data were less likely and slower to respond than researchers receiving general requests. Similarly, in 2019, a randomized controlled trial in conjunction with a Web-based survey included study authors to explore whether and how far a data-sharing agreement affected primary study authors' willingness to share IPD. [35] The response rate was relatively low (21 %) in this study since more than 1,200 individuals were initially contacted and 247 responded. Among the responders, study authors who received a data-sharing agreement were more willing to share their data set, with an estimated effect size of 0.65 (95% CI [0.39, 0.90]).

Authors of published reports on prevention or treatment trials in stroke were asked to provide data for a systematic review and randomised to receive either a short email with a protocol of the systematic review attached ('Short') or a longer email that contained detailed information, without the protocol attached ('Long').[36] 88 trials with 76 primary authors were identified in the systematic review, and of these, 36 authors were randomised to Short (trials=45) and 40 to Long (trials=43). Responses were received for 69 trials. There was no evidence of a difference in response rate between trial arms (Short vs Long, OR 1.10, 95% CI 0.36 to 3.33).

***Trial participants***

*Qualitative studies*

Perceptions of trial participants toward data-sharing and their intention to share were explored qualitatively. A systematic review with a thematic analysis of 9 qualitative studies from Africa, Asia, and North America identified four key themes emerging among patients: the benefits of data sharing (including benefit to participants or immediate community, benefits to the public and benefits to science or research), fears and harm (including fear of exploitation, stigmatization or repercussions, alongside concerns about confidentiality and misuse of data), data-sharing processes (mostly consent to the process), and the relationship between participants and research (e.g. trust in different types of research or organizations, relationships with the original research team).[37] Some qualitative reports provide data on heterogenous samples including patients and various stakeholders from low- and middle-income countries. In-depth interviews and focus group discussions involving 48 participants in Vietnam suggested that trial participants could be more willing to be involved in data-sharing than trialists.[38] A similar study on a range of relevant stakeholders in Thailand found that data-sharing was seen as something positive (e.g. a means to contribute to scientific progress, better use of resources, greater accountability, and more output) but it underlined considerable reservations, including potential harm to research participants, their communities, and the researchers themselves.[39]

In a qualitative study with 16 in-depth interviews, cancer patients currently participating in a clinical trial indicated a general willingness to allow re-use of their clinical trial data and/or samples by the original research team, and supported a generally open approach to sharing data and/or samples with other research teams, but some would like to be informed in this case.[40] Despite divergent opinions about how patients prefer to be involved, ranging from passive contributors to those explicitly wanting more control, participants expressed positive opinions toward technical solutions that allow their preferences to be taken into account.

*Surveys*

Two surveys performed in the US and one in Italy assessed the intention-to-share rates among trial participants (see **Figure 4**). In one survey with a moderate response rate (47%), 463/799 (58%) patients favored or strongly favored data-sharing, while only 9% were against or strongly against it.[41] Most participants (84%) believed that disclosing the data-sharing plan within the informed consent process was important or very important. A higher percentage of ethnic minority participants was against data-sharing (white, 6%, vs. "other", 13).

In a second survey with a high response rate (79%), 93% were very or somewhat likely to allow their own data to be shared with university scientists and less than 8% of respondents felt that the potential negative consequences of data-sharing outweighed the benefits.[42] Predictors of this outcome were a low level of trust in others, concern about the risk of re-identification or about information theft, and having a college degree. 93% and 82 % respectively were very or somewhat likely to allow their data to be shared with academic scientists and scientists in for-profit companies. The purpose for which the data would be used did not

influence willingness to share data except for use in litigation. However, patients were concerned that data-sharing might make others less willing to enroll in clinical trials, that data would be used for marketing purposes, or that data could be stolen. Less concern was expressed about discrimination and exploitation of data for profit.

In a survey of Italian patient and citizen groups, 280/2003 (14%) contacts provided questionnaires eligible for analysis.[43] 144/280 (51%) had some knowledge about the IPD sharing debate and 60/280 (42%) had an official position. Of those who had an official position 35/60 (58%) were in favour and 19/60 (32%) in favour with restrictions. 39% approved broad access by researchers and other professionals to identified information.

*Other data sources*

While consent seems to be a crucial issue for trial participants, an analysis of 98 Informed Consent Forms (ICFs) found that only 6 (4%) indicated a commitment to share de-identified IPD with third party researchers.[44] Commitments to share were more common in publicly funded trials than in industry-funded trials (7% vs 3%).

***Publishers/funders***

*Publishers*

*Metrics of data sharing statements and policies*

Several studies were found about the intentions (and data-sharing policies) of publishers. Many publishers have developed data-sharing policies (20-75%), however, less than 10% are mandatory (see **Figure 4**). In a 2009 survey of editors of different member journals of the World Association of Medical Editors (WAME) (response rate 22%), 2% and 19% of journals respectively required provision of participant level data and specification by authors of their data-sharing plan.[45] A similar survey of 10 high-impact surgical journals in 2009 and 2012 found only one journal that had a mandatory data-sharing policy.[46] Data-sharing statements were found only in 2/246 (1%) RCTs published in these 10 journals. Another study of a random sample of 60 journals found that 21 (35 %) provided instructions for patient-level data, but only 4 (7 %) required sharing of IPD (all were oncology journals).[47] A review of 88 websites of dental journals [48] suggested that 17 accepted raw data as complementary material. A 6-year cross-sectional investigation of the rates and methods of data-sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018 was performed.[49] 8/14 (57.1%) journals had data-sharing policies for published RCTs. Of the 394 RCTs included none shared their data publicly.

40/60 clinical psychology journals had a specific policy for data-sharing (2017).[50] Only one journal made data-sharing mandatory, while 37 recommended it. The findings suggest great heterogeneity in journal policies and little enforcement. Online instructions for authors from 38 high-impact addiction journals were reviewed for 6 publication procedures, including data-sharing (2018). 28/38 (74%) of the addiction journals had a data-sharing policy, none was mandatory.[51] It was concluded that many addiction journals have adopted publication policies, but more stringent requirements have not been widely adopted. Instructions for authors in 43 high-impact nutrition and dietetics journals were reviewed with respect to procedures to increase research transparency (2017).[52] 25/33 (75%) journals publishing original research and 4/10 review journals had a data-sharing policy.

Among 109 peer-reviewed and original research-oriented dental journals that were indexed in the MEDLINE and/or SCIE database in 2018, a data-sharing policy was present in 32/109 (29.4%) and 2 of these had a mandatory policy.[53] This study concluded that at present data-sharing policies are not widely endorsed by dental journals. In a cross-sectional survey 14 ICMJE-member journals and 489 ICMJE-affiliated journals that published an RCT in 2018 were evaluated with respect to data-sharing recommendations.[54] 8/14 (57%) of member journals and 145/489 (30%) of affiliated journals had an explicit data-sharing policy on their website. In RCTs published in member journals with a data-sharing policy, there were data-sharing statements in 98/100 (98 %) with expressed intention to share individual patient data in 77/100 (77%). In RCTs published in affiliated journals with an explicit data-sharing policy, data-sharing statements were rare 25/100 (25%), and expressed intentions to share individual participant data were found in 22/100 (22%).

Changes in policies from 2013 to 2016 regarding public availability of published research data were investigated in 115 paediatric journals.[55] In 2012 77 /115 (67%) and in 2016 56/115 (49%) accepted storage in thematic or institutional repositories. Publication of data on a website was accepted by 27/115 (23%) and 15/115 (13%). Most paediatric journals recommend that authors deposit their data in a repository but they do not provide clear instructions for doing so.

*Funders and clinical trial units*

*Metrics of data sharing policies by funders*

Several studies investigated mandatory data-sharing policies of funders. 30-80% of the non-commercial funders provided data-sharing policies, the highest rates were observed in the US. Only around 10-20% of these policies were mandatory (see **Figure 4**). In one study 50% of the top non-commercial funders had a data-sharing policy but it was found that in only 2/20 cases data-sharing was required. Six funders offered technical or financial resources to support IPD sharing.[56] Trial transparency policies were investigated for 9/10 top non-commercial funders in the US (May to November 2018).[57] 7/9 (78%) funders had a policy for individual patient data-sharing, for 1 it was mandatory. 6 offered data-sharing and 5 monitored compliance. Of 96 responders out of 190 non-commercial funders contacted in France, 31 were identified as funding clinical trials (2019).[58] 9/31 (29%) had implemented a data-sharing policy. Among these 9 funders, only one had a mandatory sharing policy and 8 a policy supporting but not enforcing data-sharing. Funders with a data-sharing policy were small funders in terms of total financial volume.

Three studies investigated mandatory data sharing policies among commercial sponsors (see **Figure 4**). In a 2016 survey, 22/23 (96%) companies among the top 25 companies by revenue had a policy to share IPD. In a second sample of 42 unselected companies, 30 (71 %) had one. These policies generally did not cover unlicensed products or trials for an off-label use of a licensed product. 52 % of top companies, and 38 in the sample including all companies considered that requests for IPD for additional trials were not explicitly covered by their policy.[59] A second survey studied data availability for 56 publications reporting on 61 industry-sponsored clinical trials of medications.[60] Of these 61 studies, 32 (52%) had a public data-sharing policy/process.

78 non-commercial funders and a sample of 100 leading commercial funders in terms of drug sales having funded at least one RCT in the years 2016 to 2018 were surveyed (15 February 2019 – 10 September 2019).[61] 30/78 (38%) non-commercial funders had a data-sharing policy with 18/30 (60%) making data-sharing mandatory and 12/30 (40%) encouraging data-sharing. 41/100 (41%) of the commercial funders had a data-sharing policy. Among funders with a data-sharing policy, a survey of two random samples of 100 RCTs registered on Clinicaltrial.gov found that data-sharing statements were present for 77/100 (77%) and 81/100 (81%) of RCTs funded by non-commercial and commercial funders respectively. Intention to share data was expressed in 12/100 (12%) and 59/100 (59%) of RCTs funded by non-commercial and commercial funders. The survey indicated suboptimal performance by funders in setting up data-sharing policies.

*Metrics of data-sharing policies by CTUs*

Among 23 UK Clinical Research Collaboration (UKCRC) registered Clinical Trial Units (CTUs) (response rate = 51 %), 5 (22 %) had an established data-sharing policy and 8 (35%) specifically required consent to use patient data beyond the scope of the original trial (see table).[10] Concerns were raised about patient identification, misuse of data, and financial burden. No CTUs supported the use of an open access model for data-sharing.

*Other data sources*

A 2005 survey of 107/122 accredited medical schools in the US (response rate = 88%) explored data-sharing in the context of contractual provisions that could restrict investigators' control over data in the context of industry-funded trials.[62] There was poor consensus among senior administrators in the offices of sponsored research at these institutions on the question of prohibiting investigators from sharing data with third parties after the trial is over (41 % allowed it, 34 % disallowed it, and 24 % were not sure whether they should allow it).

In a survey targeting European heads of imaging departments and speakers at the clinical trials in radiology sessions (July – September 2018), the response rate was 132/460 (29%).[63] Responses were received from institutions in 29 countries, reporting 429 clinical trials. For future trials, 98% of respondents (93/95) said they would be interested in sharing data, although only 34% had already shared data (23/68). The main barriers to data-sharing were data protection, ethical issues, and lack of a data-sharing platform.

**Results for individual sources of evidence: actual data-sharing**

***Re-users***

*Studies related to journal articles*

*Metrics of actual data-sharing*

Several studies have been performed investigating data-sharing rates for studies that have been published in journals, the majority with data-sharing policies and high impact (**Figure 5**). Even with strict data-sharing policies, the data-sharing rates are low or at most moderate, and vary between 10 and 46%, except for one study with a very high data-sharing rate due to a partly preselected sample of authors willing to share their data [19]. In the 6-year cross-sectional investigation of the rates and methods of data-sharing in 15 high-impact addiction journals that published clinical trials between 2013 and 2018, none of the 394 clinical trials included shared their data publicly [49]. Of 86 responders in a survey targeting the corresponding authors of 619 RCTs published between 2014 - 2016 in 7 high-ranking anaesthesiology journals, raw data was obtained only for 24 studies.[25] 62 declined to share raw data. In a study targeting PLOS Medicine and PLOS Clinical Trials publications conducted in 2009, 1/10 (10%) of the data sets was made available after request [64]. In articles in Chinese and international journals from 2016, sharing practices were indicated for 29/247 (11%) of the articles.[20] Among the top 10 general and internal medical journals investigated in 2016, IPD was provided after request for 9/61 (15%) of pharmaceutical-sponsored studies [60]. For BMJ research articles published between 2009 and 2015, data sets were made available in 7/157 (4%) of the articles[23]. For the sub-sample of clinical trials, the rate was higher (5/21 (24%)). Of 317 clinical trials published in 6 general medical journals between 2011 and 2012, 115 (36%) granted access to data[17]. The data availability for RCTs published in BMJ and PLOS Medicine between 2013 and 2016 was 17/37 (46%)[65].

*Experimental studies*

In a parallel group RCT, an intervention group (offer of an Open Data Badge for data-sharing) was compared to a control group (no badge for data-sharing).[66] The primary outcome was the data-sharing rate. Of 160 research articles published in BMJ Open, 80 were randomised to the intervention and control groups, of which 57 could be analysed in the intervention group and 54 in the control group. In the intervention group data was available on a third-party repository for 2/57 (3.5%) and upon request for 32/57 (56.1%) respectively in the control group: 3/54 (5.6%) and 30/54 (56%). Data-sharing rates were low in both groups and did not differ between groups.

*Data sharing for IPD meta-analyses*

*Metrics of data-sharing for IPD meta-analyses*

Some examples demonstrate that data availability for IPD meta-analyses is still limited despite the various data-sharing initiatives/platforms (**Figure 5**). Availability can be increased under specific circumstances, such as the creation of a disease-specific repository for a scientific community, as demonstrated for a repository of IPD from multiple low back pain RCTs with IPD from 20/42 (48%) RCTs included [67] and a study on anti-epileptic drugs conducted by a Cochrane group with IPD for 15/39 (38%) studies included [68]. In another study on different databases, 35 individual participant data meta-analyses with more than 10 eligible RCTs were identified (May 1, 2015 to February 13, 2017)[69]. Of 774 eligible RCTs identified in these meta-analyses, 517 (66.8 %) contributed data. The country where RCTs are conducted (the UK versus the United States (US)), the impact factor of the journal (high versus low) and a recent RCT publication year were associated with higher sharing rates. In three other studies, the availability of datasets for IPD meta-analysis was limited (0-17%). In one study performed in 2014, devoted to one commercial sponsor with one specific medicinal product, IPD from 24 trials was requested without success [70]. Of 15 requests (13 direct to authors, 2 to a repository) in 2014/2016, IPD was received for 2/15 (13%) of the studies [71]. Of 217 RCTs published since 2000 in orthopaedic surgery, agreement to send IPD was obtained for 37/217 (17%)[72].

*Experimental studies*

The low data availability for IPD-meta-analyses is underlined by two experimental studies. One experimental study covered the issue of actual data-sharing. In this small randomized prospective study where 29 corresponding authors of original research articles in a medical journal were contacted via two different modes (general versus specific request), only one author actually sent the data immediately in response to a specific request and one author, without caveats, reported willingness to send the data in response to a general request.[34]

A randomized controlled trial investigated the effect of financial incentives on IPD sharing.[73] All study participants (129 in all) were asked to provide the IPD from their RCT. Those allocated to the intervention group received financial incentives, those from the control group did not. The primary outcome was the proportion of authors who provided IPD. None of the authors shared their IPD, whichever the group.

*Other data sources*

Two studies investigated the completeness of data availability in IPD meta-analyses. Out of 30 IPD meta-analyses included in a survey,[74] 16 did not have all the IPD data requested. The access rate for retrieving IPD for use in IPD-meta-analyses was investigated in a systematic review.[68] Only 188 (25%) of 760 IPD meta-analyses retrieved 100% of the eligible IPDs for analysis and there was poor evidence that IPD retrieval rates improved over time.

*Access to repositories/platforms*

Only a few studies describe access to repositories/platforms from the viewpoint of the user (**Figure 5**). Experiences with two major platforms (CSDR, PDS) were reported.[75] In these very early-phase projects, no data access was possible with CSDR, and faster data acquisition was achieved via the Project Data Sphere. High sharing rates were reported for academic repositories (MRC CTU, BioLINCC). Of 103 requests to MRC CTUs, access was granted in 80/103 (78%) cases [76]. In a survey of investigators 536/536 (100%) received access to BioLINCC over a time period between 2007 and 2014 [77].

***Repositories/platforms***

*Commercial sponsors*

*Metrics of actual re-use*

Different initiatives and platforms were initially implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now open to academic trials but this has not been used very often so far). This covers the YODA project, CSDR, Vivli and SOAR (which is now part of Vivli). For the different platforms and repositories, metrics describing the actual use of data are available (**Figure 5**).

6 studies have accessed data-sharing rates for CSDR. From 2014 to the end of January 2019, there was a total of 473 research proposals submitted to CSDR.[78] Of these, 364 met initial administrative and data availability checks, and the independent review panel approved 291. 222/473 (46.9%) of the requests gained access to the data (in progress and completed). Of the 90 research teams that had completed their analyses by January 2018, 41 reported at least one resulting publication to CSDR. Less than half of the studies ever listed on CSDR have been requested. Between 2014 and 2017 CSDR received a total of 172 research proposals, of which 105 (61%) were approved [79]. In another study focusing on availability and use of shared data from cardiometabolic clinical trials in CSDR covering the time period between 2013 and 2017, 198 (62%) were approved with or without conditions [80]. In year one of the use of CSDR (2013-2014), 36 research proposals were approved with conditions, of these 23 (64%) progressed to a signed data-sharing agreement [81].From 2014 to 2017, Boehringer-Ingelheim listed 350 trials for potential data-sharing at CSDR.[82] 55 research proposals were submitted, of which 37 (67.3%) were approved. All approved research proposals submitted to Boehringer-Ingelheim except one addressed new scientific questions or were structured to generate new hypotheses for further confirmatory research, rather than replicating analyses by the sponsor to confirm previous research.

Between 2013 and 2015, 177 research proposals were submitted to CSDR, and access was granted for 144 (81%) of these proposals [83].

In the first year following the launch in October 2014, YODA received 29 requests all of which were approved (100%) [84]. In 2017 the YODA project reported 73 proposals of which 65 were approved [85]. A more recent publication reported the metrics for data-sharing of Johnson & Johnson clinical trials in the YODA project up to August 27, 2018.[86] 100 data requests were received from 89 principal investigators (PI) for a median of 3 trials per request. 90/100 requests (90 %) were approved and a data use agreement was signed in 82/100 (82%).

The use of the open access platforms CSDR, YODA and SOAR together between 2013 and 2015 was investigated in one study. Of the 234 proposals submitted, 154 (66%) were approved [87].

The data available shows that the use of these platforms has increased steadily since their initiation and that 50% and more of the data requests lead to actual data-sharing. The reasons for not sharing are numerous but data access is rarely denied by the platforms. Our assessment of CSDR, YODA, NIDDK and Vivli websites is presented in **Table 1**.

**Table 1**: Metrics of CSDR, YODA, and Vivli websites
NIDDK also provided metrics concerning the number of requests (530) but no other information
*publication anticipated

| Platform | Metrics date | Available studies | Number of requests | Number of requests with data shared | Number of requests with data leading to publication | Number of publications |
|---|---|---|---|---|---|---|
| CSDR | 30/11/2020 | 3008 | 621 | 318 | 59* | 79 |
| YODA | 15/11/2019 | 334 | 196 | 173 | 29 | 35 |
| Vivli | 02/11/2020 | 5203 | 215 | 123 | 8 | 9 |

*Metrics of trial coverage for data-sharing*

Ethics approval in applications for open-access clinical trial data from CSDR was investigated in a survey.[79] Projects with and without ethics approval were applied to at roughly similar rates (62/111 and 43/61).
The proportion of trials where the pharmaceutical and medical device industry provided IPD for secondary analyses and thus the completeness of trial data is still limited.[60] Only 15% of 61 industry-sponsored clinical trials were available 2 years after publication. For companies listing at least 100 studies on CSDR, a search was performed in ClinicalTrials. gov (January 2016, studies terminated/ completed at least 18 months before search date).[88] Among 966 RCTs registered in ClinicalTrials.gov, only 512 (53%) were available on CSDR and only 385 (40%) of the RCTs were registered and listed on CSDR with all datasets and documents available. This was the case despite the time lapse of 18 months since the completion of the drug trials by the company sponsor. Differences across sponsors were observed. Pharmaceutical repositories may cover only part of the trials with commercial sponsors needed for meta-analyses. In a study investigating data availability for industry-sponsored cardiovascular RCTs with more than 5000 patients, performed by a top-20 pharmaceutical company and registered at ClinicalTrials.gov (up to Jan. 2015), only 25% of the identified trial data was confirmed to be available.[24] In 50% of cases availability could not be definitely confirmed.
As part of the Good Pharma Scorecard project, data-sharing practices were assessed for large pharmaceutical companies with novel drugs approved by the FDA in 2015, using data from ClinicalTrials.gov, Drugs@FDA, corporate websites, data-sharing platforms and registries (e.g. YODA, CSDR)[89]. 628 trials were analysed. 25% of the large pharmaceutical companies made IPD accessible to external investigators for new drug approvals, this proportion improved to 33% after applying a ranking tool.

*Non-commercial sponsors*

Disease-specific academic clinical trial networks have a long history of IPD sharing, especially US-related NIH institutions. This is clearly demonstrated by the available literature; however, the metrics of data-sharing are

not always as transparent as with the industry platforms, and data cannot be structured and documented easily in a table.

In a survey on the use of the National Heart, Lung, and Blood institute Data Repository, access to 100 studies initiated between 1972 and 2010 was investigated.[90] A total of 88 trial datasets were requested at least once, and the median time from repository availability and the first request was 235 days.

Since its inception in 2006 and through to October 2012, nearly 1700 downloads from 27 clinical trials have been accessed from the Data Share website belonging to the National Drug Abuse Treatment Clinical Trial Network (CTN) in the US, with use increasing over the years.[91] Individuals from 31 countries have downloaded data so far.

In a case study approach, the data-sharing platform Data Share of the National Institute of Drug Abuse (NIDA) was investigated in detail.[92] As of March 2017, the Data Share platform had included 51 studies from two trial networks (36 studies from CTN and 15 studies from NID Division of Therapeutics and Medical Consequences). From 2006 to March 2017, there have been 5663 downloads from the Data Share website. Of these, 4111 downloads have been from the US.

The Project Data Sphere (PDS) is an open-source data-sharing model that was launched in 2014 as an independent, non-profit initiative of the CEO roundtable on cancer.[93] PDS contains data from 72 oncology trials, donated by academics, governments, and industry sponsors. More than 1400 researchers have accessed the PDS database more than 6500 times. As an example, a challenge to create a better prognostic model for advanced prostate cancer was issued in 2014, with 549 registrants from 58 teams and 21 countries.

The Immune Tolerance Network (ITN) is a National Institute of Allergy and Infectious Diseases /National Institutes of Health-sponsored academic clinical trial network.[94] The trial sharing portal, which was released for public access in 2013, provides complete open access to clinical trial data and laboratory studies from ITN trials at the time of the primary study publication. Currently, data from 20 clinical trials is available and data for an additional 17 will be released to the public at the timepoint of first publication. So far, more than 1000 downloads have been registered.

In the MRC Clinical Trials Transparency Review Final Report (November 2017), the MRC United Kingdom (UK) reported that 24/107 (22%) trials that started during the review period had created a database for sharing. Seven of these datasets (7/24, 29%) had already been shared with other researchers.[95]

Of 215 requests submitted for PLCO (Prostate, Lung, Colorectal and Ovarian) cancer screening trial data, 199 (93%) were approved, and for NLST (National Lung Screening Trial) 214 (89%) out of 240 requests.[96]

*Other stakeholders*

In a case study about experiences with data-sharing among data monitoring committees, access to five concurrent trials assessing the level of arterial oxygen, which should be targeted in the care of very premature neonates, was investigated.[97] The target of taking all relevant evidence into account when monitoring clinical trials could be only partially reached.

One case-study directly addressed the issue of costs. Data from two UK publicly funded trials was used to assess the resource implications of preparing IPD from a clinical trial to share with external researchers.[98] One trial, published in 2007, required 50 hours of staff time with a total estimated cost of £3185, and the other published in 2012 required 39.5 hours with £2540.

**Results of individual sources of evidence: re-use**

***Any type of re-use***

The majority of research projects using shared clinical trial data are dealing with new research. This covers studies on risk factors and biomarkers, methodological studies, studies on optimizing treatment and patient stratification and subgroup analyses. IPD meta-analyses were a less frequent reason for data-sharing requests to repositories and only a few have been reported. Re-analyses are only exceptionally applied.

Early experiences with CSDR, involving GlaxoSmithKline trials found low rates of IPD meta-analyses and re-analyses, the vast majority being secondary analyses (studies on risk factors or biomarkers, methodological studies, predictive toxicology or risk models, studies of optimizing treatments, subgroup analyses etc.).[81] Similar results were found in an update of the analysis.[83]

In the YODA project, which had received 73 proposals for data-sharing as of June 2017 and had approved 65 proposals, the most common study purposes were to address secondary research questions (n=39), to combine data as part of larger meta-analyses (n=35) and/or to validate previously published studies (n=17) [85].

Among the 172 requests to the National Heart, Lung and Blood Institute (NHLBI) data repository with online project descriptions and coded purpose, 72% of requests were initiated to address a new question or hypothesis, 7% to perform a meta-analysis or combined study analysis, 2% to test statistical methods, 9% to investigate methods relevant to clinical trials, and 9% for other reasons.[90] In only two requests, the available description suggested a re-analysis.

From 2014 to the end of January 2019, 222/473 (46.9%) of the requests to CSDR gained access to the data (in progress and completed).[78] 90/222 (40.5 %) of the research teams had completed their analyses by January 2018. 41 published at least one paper, and another 28 that were expected to publish shortly.

In the SPRINT challenge, individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and to identify novel scientific or clinical findings.[99] Among 200 qualifying teams, 143 entries were received.

***Further additional analyses***

There were few indications concerning the exact type of secondary analysis that was performed. Approved proposals per subject matter are available for the Cancer Data Access system (CDAS), covering two large cancer screening trials (PLCO, NLST).[96] Of the 199 approved requests to PLCO between November 2012 and October 2016, 84 (42%) were devoted to cancer etiology, 66 (33%) to trial-related screening, 29 (15%) to other areas, 14 (7%) to risk prediction and 6 (3%) to image analysis. Of the 214 approved requests to NLST, 95 (44%) were devoted to image analysis, 90 (42%) to trial-related screening, 14 (7%) to other subjects, 10 (5%) to cancer etiology and 5 (2%) to risk prediction.

***IPD meta-analyses***

In one study, IPD meta-analyses proved to amount to a small proportion of data re-use. Among the 174 research proposals approved up to 31 August 2017 by CSDR, 12 proposals were IPD meta-analyses, including network meta-analyses.[71] All were retrospective IPD meta-analyses (i.e. none was a prospective IPD meta-analysis).

***Re-analyses***

A 2014 survey of published re-analyses [100] found that a small number of reanalyses of RCTs have been published (only 37 re-analyses of 36 initial RCTs) and only a few were conducted by entirely independent authors. 35% of these reanalyses led to changes in findings that implied conclusions different from those of the original article for the types and numbers of patients who should be treated.

In the survey of 37 RCTs in the BMJ and PLOS Medicine published between 2013 and 2016, 14 out of 17 (82%, 95% IC: 59% to 94%) available studies were fully reproduced on all their primary outcomes.[65] Of the remaining RCTs, errors were identified in two, but reached similar conclusions, and one paper did not provide enough information in the Methods section to reproduce the analyses.

**Results for individual sources of evidence: output from data sharing**

Publications can be considered as the main research output of data-sharing. Publication activity in the re-use of clinical trial data was considered in several studies. Detailed data are available for academic clinical trial networks and disease-specific repositories in the US, some of them already practising data-sharing for a period longer than 10 years. Here, fair to moderate publication output has been observed depending on the individual repository. So far this is not the case for the repositories storing clinical trial data from commercial sponsors, taking into consideration that these repositories were established around five years ago and that there is

usually a considerable time lag between request, approval, analysis and publication. Current statistics indicate improvement in publication output with time.

***Non-commercial sponsors***

In a cross-sectional web-based survey about access to clinical research data from BioLINCC, covering the period from 2007 to 2014, 98 out of 195 responders (50%) reported that their projects had been completed, among which 66 (67%) had been published.[77] Of the 97 respondents who had not yet completed their proposed projects, 81 (84%) explained that they planned to complete their project; 63 (65%) indicated that their project was in the analysis/manuscript draft phase.

In a survey targeting European heads of imaging departments and speakers at the Clinical Trials in Radiology sessions (July – September 2018), 23/68 reported that they had already shared data.[63] At least 44 original studies were published based on the data shared by the 23 institutions involved.

In five studies (**Table 2**) the number of publications was reported, usually referring to the number of trials included in the repository/platform.

**Table 2**: Studies reporting published outputs for non-commercial sponsors

| Reference | Repository/ platform | No. of trials included in repository/platform | No. of published articles | Assessment |
|---|---|---|---|---|
| Shmueli-Blumberg, 2013 | CTN Data Share | 27 trials (1700 downloads) | 13 | 2012 |
| Zhu, 2017 | CDAS | 2 trials (PLCO, NLST) (455 requests) | 25% for PLCO projects, 19% for NLST projects | 2016 |
| Coady, 2017 | BioLINCC | 100 trials (88 requested at least once) | 35% of clinical trials at least 1 publication 5 years after availability in the repository | 5/2016 |
| Huser, 2018 | NIDA Data Store | 51 trials | 14 | 3/2017 |
| Pisani, 2017 | WWARN | 186 trials | 18 | 2016 |

***Commercial sponsors***

Various studies explored metrics of both YODA and CSDR (**Supplementary Material 4**).

Up to 2021, Vivli's website indicates very little published output. We were not able to retrieve published output from NIDDK. **Figure 6** presents publication metrics for CSDR (up to 31 August 2019) and YODA (up to 1st July 2019). Among 88 published papers (62 from CSDR and 26 from YODA), 49 were secondary analyses (42 from CSDR and 7 from YODA), 30 were meta-analyses (13 from CSDR and 17 from YODA), 6 were methodological studies (5 from CSDR and 1 from YODA) and 3 were re-analyses (2 from CSDR and 1 from YODA). The details of these publications are presented in **Supplementary Material 5**. [80 83 85]

**Results of individual sources of evidence: impact of research output**

Evidence on the impact of research output from sharing IPD from clinical trials is still very sparse. So far only two studies, with inconsistent results dealing with this issue and focusing only on citation metrics could be identified.

*Metrics on citations*

One study, already published in 2007, suggested that sharing detailed research data was associated with an increased citation rate.[101] Of 85 cancer microarray clinical trials published between January 1999 and April 2003 41 made their microarray data publicly available on the internet. For 2004 – 2005, the trials with publicly available data received 85% of the aggregate citations. Publicly available data was significantly associated with a 69% increase in citations, independently from journal impact factor, date of publication and the author's country of origin.

Citation metrics for 224 publications based on repository data for clinical trials in the NHLBI Data Repository were compared with publications that used repository observational study data, as well as a 10%-random sample of all NHLBI-supported articles published in the same period (January 2000 – May 2015).[90] Half of the publications based on clinical trial data had cumulative citations that ranked in the top 34% normalized for subject category and year of publication, compared to 28.3% for publications based on observational studies and 29% for random samples. The differences were, however, not statistically significant.

*Other data sources*

In the SPRINT challenge, individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and to identify novel scientific or clinical findings.[99] Among 200 qualifying teams, 143 entries were received. Entries were judged by a panel of experts on the basis of the utility of the findings to clinical medicine, the originality and novelty of the findings, and the quality and clarity of the methods used. All submissions were also open for crowd voting among the 16,000 individuals following the SPRINT Challenge. Cash prizes were awarded, and winners were invited to present their results. 143 entries to the SPRINT data challenge were received.

## DISCUSSION

### Summary of evidence

There are major differences with respect to the intention to share IPD from clinical trials across the different stakeholder groups. The studies available so far show that clinical trialists and to some extent study participants, as the two main actors of clinical trials, usually have great willingness to share data (60-80%). This is much less pronounced when it comes to data-sharing statements published in journal articles. Depending on the journals considered, the rates vary from less than 5% to around 25%. The situation is even worse when data-sharing plans documented in registries (e.g. ClinicalTrials.gov) are analysed. Here the willingness to share data is between 5 and 10%.
As a consequence, considerable discrepancy between the positive attitude towards data-sharing in general and the intention to do so in an actual study needs to be ascertained. Publishers, enabling the publication of research output from clinical trials and funders/sponsors financing clinical trials, could be major drivers to change the situation. Meanwhile many publishers have developed data-sharing policies (20-75%), but less than 10% are mandatory and have thus not been enforced. There are differences between journals, with some of the high-impact journals being more involved in the data sharing movement than the others *(e.g PLOS Medicine, the BMJ, Annals of Internal Medicine).* For funders, the situation is similar, but differs between commercial and non-commercial funders. 30-80% of the non-commercial funders provide data-sharing policies, with the US and NIH at the front. Only around 10 to 20% of these policies are mandatory. Data-sharing policies have been developed more often in the group of commercial funders (40-95%) but information on the proportion of mandatory policies is lacking. In short, the pressure by publishers and funders to share data is still limited and the situation is only slowly improving. Stronger policies on data sharing that include a strong evaluation component are needed. The situation is better for the pharmaceutical industry, which has not only promoted data-sharing policies in their organisations to a large degree but has also implemented platforms and repositories, providing practical support for the process of data-sharing (e.g. CSDR, Yoda, Vivli).

Several studies have been performed investigating data-sharing rates for clinical studies that have been published in journals. The focus has been on high-impact journals with strict data-sharing policies (e.g. PLOS Medicine, BMJ, Annals of Internal Medicine), demonstrating data-sharing rates between 10% and 46%, except for one study with a very high data-sharing rate due to a partly preselected sample of authors willing to share

their data. Data availability for IPD meta-analyses is usually limited (0-20%), available only under specific circumstances (Cochrane group, disease-specific repository) and the availability can be increased to 50% and more. A few individual studies describe access to repositories/platforms from the viewpoint of the user, which does not enable identification of a general pattern. Different initiatives and platforms have been implemented for the pharmaceutical and medical device industry to support sharing of IPD from clinical trials (these platforms are now open to academic trials, but this has not been used very often so far). This covers the YODA project, CSDR, Vivli and SOAR (which is now part of Vivli). The data available shows that the use of these platforms has increased steadily since their initiation and that 50% and more of the data requests lead to actual data-sharing. The reasons for not sharing are numerous but data access is rarely denied by the platforms. One of the hurdles to better acceptance of data sharing is the time delay between a request for data sharing and receiving the requested data. This was not systematically investigated in the scoping review, but a few studies have demonstrated that there may be a considerable time lag between initial request and response [68 73] and the time between request and receiving a data sharing agreement [75].

The majority of research projects using shared clinical trial data deal with new research. This covers studies on risk factors and biomarkers, methodological studies, studies on optimizing treatment and patient stratification and subgroup analyses. This is important because new research may be easier to publish in peer-reviewed journals, which is a major driver of academic careers.

So far only some IPD meta-analyses have been planned as part of data-sharing initiatives, and only a few have been reported. There are many hurdles for IPD meta-analyses, including the findability, the accessibility and the re-usability of datasets (F, A and R in FAIR). ECRIN has developed a metadata dictionary (MDR), able to identify clinical studies and data objects related to it (e.g. protocol, DMP, CRF).[102] This tool allow for identifying studies for which datasets are available and the conditions for access (ECRIN, MDR). Even if IPD datasets are accessible for meta-analyses, the studies are usually distributed across various repositories. This has been demonstrated in several studies in our scoping review. One central repository could simplify the situation, but instead, the number of repositories is steadily increasing.[2] The situation could be considerably improved with more standardisation and harmonisation of data and procedures and a federating approach between repositories.

Re-analysis of clinical trial data could help the scientific community to enhance the validity of reported trial results. An illustration is the “restoring study 329” initiative, investigating efficacy and harm of paroxetine and imipramine in the treatment of major depression in adolescence. The re-analysis reached different conclusions with important implications for both clinical practice and research.[3] RIAT (Restoring invisible & abandoned trials support center) was initiated as an international effort to tackle bias in the way research is reported with the goal of providing more accurate information to patients and other healthcare decision makers.[103]

One of the problems that is tackled by RIAT is misreporting (inaccurately or incompletely reported trials). In our scoping review we found that re-analyses are only exceptionally applied. In one review, the majority of studies was reproduced on all primary outcomes, in another around one third of studies led to changes in findings different from the original articles. It seems that re-analysis is only attractive in a minority of cases deserving major public interest. Nevertheless, for these cases, repositories holding and sharing IPD could be very useful and speed up the process of data-sharing. It could be of interest to establish a link between RIAT and data-sharing platforms and initiatives.

Publications can be considered as the main output from data-sharing. Usually, there is a considerable time lag between requesting data for re-use, receiving shared data, performing secondary analysis, writing a manuscript and publishing the secondary analysis. This has to be taken into consideration when the publication output of data-sharing initiatives and platforms is analysed. Repositories and platforms mainly devoted to commercial trials have now existed for around 5 years, so only a limited publication output can be expected. Fortunately, these repositories provide detailed metrics for data-sharing requests, including number and type of publications originating from data-sharing. As expected, the number of publications related to data-sharing for commercial studies is still limited, but current statistics indicate improvement over time. The situation with non-commercial sponsors is different. Academic clinical trial networks and disease-repositories have been successfully implemented (mainly in the US) and have already practised data-sharing for quite a long time, some for more than 10 years. Here data-sharing is part of the research culture and the exchange of data is based on elements such as trust, technical support and common benefit. Outstanding examples are BioLINCC,[77] NIDA [91] and World Wide Antimalarial resistance Network (WWARN).[104 105] This is reflected in the data-sharing

rates for IPD meta-analyses, which are rather low if data requests target authors directly, compared to data-sharing requests within communities (e.g. Cochrane groups) or related to specific repositories. Outside clinical trial networks and disease-specific repositories, data-sharing of IPD is still very limited. Possible reasons could include the lack of widely accepted repositories for non-commercial clinical trials and insufficient incentives and benefits related to data-sharing. Some investigators may be reluctant to share their data, other may simply not know how to proceed.

We describe secondary analyses as a very popular type of reuse. These analyses are however exploratory and carry a risk of alpha inflation (due to multiple comparisons). Not all results of these analyses have been published. Alpha inflation and selective reporting can be fertile ground for non-reproducible science and this phenomenon surely deserves attention. Improvements could be achieved with a prospective registration of any protocol for secondary data use similar to the trial registries (e.g. ClinicalTrials.gov), a mandatory link between the registration and the original publication or data set and the need to refer to the primary publication or dataset if the re-analysis is published. Existing approaches and tools could then be extended to automatically identify publications related to re-use of data and establish a link to the original work (e.g. see crossmark – crossref [106] , metadata repository (MDR) developed by ECRIN linking clinical studies with related data objects).[102] Another possibility could be to set up a register for secondary analyses.

To be widely accepted, research output from shared data should have an impact on medical research (e.g. generation of new hypotheses) and medical health (e.g. changes in treatment via guidelines). Many interventions seek to maximise the benefit of trial data sharing (e.g. use of incentives for clinical trial data sharing, development of infrastructure for data sharing, etc.) but it is paramount that these interventions are evidence based. It is well known that the impact of primary studies on medical research and health often has a considerable time-lag and direct effects are not easy to demonstrate. So it is to be expected that evidence from research output from shared data is even more difficult to demonstrate. In this scoping review, taking into consideration the limited time available for data-sharing activities to generate an impact, no major effects were to be expected. As a consequence, the evidence on the impact of data-sharing is still very sparse. This could mean that it is still too early to measure any impact, or that the impact is very limited. So far, only surrogate measures have been considered (citation metrics) with inconclusive results. It is hoped that in the coming years, more studies with more relevant criteria and metrics will be performed. One option could be to closely follow up the SPRINT challenge, where 143 secondary analyses on a single clinical trial were performed, and it would be interesting to see whether one or more of these secondary analyses really had an impact.

**Limitations**

Retrieving and synthesizing information for this study proved to be difficult because we operated in a very siloed landscape where each initiative platform operates with its own metrics. We have tried to be exhaustive by reviewing both the literature and the most important initiatives. However, it was hard to keep the review up-to date as we were studying a moving target in a rapidly changing environment with more and more new initiatives. Some pharmaceutical companies may operate in their own environment and not on larger data-sharing platforms. This makes these activities even more difficult to track. In addition, data-sharing has not had a long history and many of the initiatives and activities were launched in the recent past. Therefore, only a limited research output from data-sharing can be expected so far and indeed, the number of publications is disappointing. It is expected that the number of publications will increase, and indeed we are already seeing this.

**Conclusions**

There is currently a gap in the evidence base evaluating impact of IPD sharing, which causes uncertainties in the implementation and adoption of current data-sharing policies. Data-sharing faces many challenges including, for instance, the scepticism of trialists.[107] There is therefore a need to provide high-level evidence that the value of medical research liable to inform clinical practice increases with greater transparency, and with the opportunity for external researchers to re-analyse, synthesize, or build on previous data. First, a register (such as PROSPERO[108]) for any secondary use of shared data should be created. The inclusion in such a register could be mandatory for any data-sharing agreement/publication, as for the registration of clinical trials. This register would make it possible to build an observatory of data-sharing practices providing direct feedback, without the

present silos we have to face. In addition, a register of this sort could help to prevent any selective publication of secondary analyses. Lastly, we suggest that interventional studies should be run to determine the optimal data-sharing policy and/or incentives that add value to clinical research. We do however need to take into consideration that the experimental studies performed so far were not very conclusive, indicating that experimental studies in this area are very demanding.

**ABBREVIATIONS**

ANR: Agence Nationale de la Recherche
ASAPbio: Accelerating Science and Publication in biology
BioLINCC: Biological Specimen and Data Repository Information Coordination Center
CDAS: Cancer Data Access System
CIHR: Canadian Institutes of Health Research
CSDR: Clinical Study Data Request
CTN: Clinical Trials Network
DFG: Deutsche Forschungsgemeinschaft
DGOS : Direction Générale de l'Offre de Soins
Drum: Data Repository for University of Minnesota
EBCTG: Early Breast Cancer Trialists' Collaborative Group
EC Europe: European Commission
EFPIA: European Federation of Pharmaceutical Industries and Associations
F1000Research: Faculty of 1000 Research
FAIR: Findable, Accessible, Interoperable and Reusable
ICMJE: International Committee of Medical Journal Editors
ICPSR: Inter-university Consortium for Political and Social Research
IOM: Institute Of Medicine
IPD: Individual Participant Data
ITN Trialshare: Immune Tolerance Network TrialShare
MMMP: Melanoma Molecular Map Project
MRC UK: Medical Research Council
NHMRC: National Health and Medical Research Council
NIDA: National Institute on Drug Abuse
NIDDK: National institute of Diabetes and Digestive and Kidney Diseases
NIH: National Institute of Health
NIH BioLINCC: National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center
NIHR: National Institute of Health Research
NIMH NDCT: National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness
NSFC: National Natural Science Foundation of China
PCORNeT: The National Patient-centered Clinical research Network
PHRC: Le programme hospitalier de recherche clinique
PHRMA: Pharmaceutical Research and Manufacturers of America
PI: Principal Investigator
PLOS: Public Library Of Science
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: extension for Scoping Reviews
ProAct: Pooled Resource Open Access Clinical trials database
RCT: Randomised clinical trial
RDA: Research Data Alliance
SOAR: the Supporting Open Access for Researchers initiative
SND: Swedish National Data Service
TBI-IMPACT: Traumatic Brain Injury– International Mission for Prognosis and Analysis of Clinical trials in Traumatic brain injury
The BMJ: The British Medical Journal
UK: United Kingdom
UKCRC: UK Clinical Research Collaboration
UMIN: University Medical Hospital Information Network
US: United States of America
US DoD: United States Department of Defense
Vivli: adapted from the Greek "Vivliothiki" (library) and the Latin root "viv" (life)
WWARN: World Wide Antimalarial Resistance Network
YODA: the Yale University Open Data Access Project

**FIGURES**

**Figure 1**: PRISMA flow diagram (PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses).
* For National Institute of Health (NIH US), the answer we received was not informative

**Figure 2**: Proportion of the 93 references exploring each outcome domain
Study designs considered:
- Experimental: prospective research that implies testing the impact a strategy (e.g. randomised controlled trial)
- Survey: a general overview, exploration, or description of individuals and/or research objects;
- Metrics: descriptive metrics from each initiative provided by the initiative;
- Qualitative: research that relies on non-numerical data to understand concepts, opinions or experiences.
- Other: any other research not covered above (e.g. case studies, environmental scans, etc.)"

**Figure 3**: Outcomes used to assess current data-sharing practices for individual patient data for clinical trials organized per outcome domain and number of studies exploring these outcomes.
Study designs considered:
- Experimental: prospective research that implies testing the impact a strategy (e.g. randomised controlled trial)
- Survey: a general view, exploration, or description of individuals and/or research objects;
- Metrics: descriptive metrics from each initiative provided by the initiative;
- Qualitative: research that relies on non-numerical data to understand concepts, opinions or experiences.
- Other: any other research not covered above (e.g. case studies, environmental scans, etc.)"

**Figure 4**: Intent to share

Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference

a: The proportion is 73 % if the purpose is a re-analysis

b: 54 participants out of 60 had an opinion about data-sharing (the others had no knowledge or no opinion)

c: An additional 25 % were undecided

d: The proportion is 19 % for requiring a data-sharing plan

e: 35 % have a data-sharing policy (encouraging data-sharing)

f: Only 2 with a mandatory policy

g: The proportion is 71% for a sample of all companies (not only the top 25)

In DeVito et al. 2018, we extracted the information on policies that made data-sharing mandatory (i.e. a requirement to share the data).

**Figure 5**: Actual data-sharing

Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference

**Figure 6**: Temporal trends, number and type of published output from CSDR and YODA
- Blue: YODA
- Red: CSDR

1. Nosek BA, Alter G, Banks GC, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. *Science (New York, NY)* 2015;348(6242):1422-5. doi: 10.1126/science.aab2374 [published Online First: 2015/06/27]
2. Banzi R, Canham S, Kuchinke W, et al. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials* 2019;20(1):169. doi: 10.1186/s13063-019-3253-3 [published Online First: 2019/03/17]
3. Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. [published Online First: 2015/09/18]
4. Data sharing and the future of science. *Nature Communications* 2018;9(1):2817. doi: 10.1038/s41467-018-05227-z
5. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data: A Proposal from the International Committee of Medical Journal Editors. *PLoS Med* 2016;13(1):e1001950. doi: 10.1371/journal.pmed.1001950 [doi]
PMEDICINE-D-15-03676 [pii] [published Online First: 2016/01/21]
6. Rosenbaum L. Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp. *N Engl J Med* 2017;376(23):2201-03. doi: 10.1056/NEJMp1704482 [doi] [published Online First: 2017/04/27]
7. Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *PLoS Med* 2017;14(6):e1002315. doi: 10.1371/journal.pmed.1002315 [doi]
PMEDICINE-D-17-01550 [pii] [published Online First: 2017/06/06]
8. Vasilevsky NA, Minnier J, Haendel MA, et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* 2017;5:e3208. doi: 10.7717/peerj.3208 [published Online First: 2017/05/04]
9. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ open* 2017;7(12):e018647. doi: 10.1136/bmjopen-2017-018647 [published Online First: 2017/12/17]
10. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. *Journal of clinical epidemiology* 2016;70:17-25. doi: 10.1016/j.jclinepi.2015.07.002 [published Online First: 2015/07/15]
11. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 2016;3:160018. doi: 10.1038/sdata.2016.18 [published Online First: 2016/03/16]
12. Peters MD, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare* 2015;13(3):141-6. doi: 10.1097/xeb.0000000000000050 [published Online First: 2015/07/03]
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of internal medicine* 2018;169(7):467-73. doi: 10.7326/m18-0850 [published Online First: 2018/09/05]
14. Clinicaltrials.gov. Learn About Clinical Studies [Available from: https://clinicaltrials.gov/ct2/about-studies/learn accessed 2019, november 21th.
15. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of clinical epidemiology* 2016;75:40-6. doi: 10.1016/j.jclinepi.2016.01.021 [published Online First: 2016/03/24]
16. Tudur Smith C, Dwan K, Altman DG, et al. Sharing Individual Participant Data from Clinical Trials: An Opinion Survey Regarding the Establishment of a Central Repository. *PloS one* 2014;9(5):e97886. doi: 10.1371/journal.pone.0097886

17. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. *Bmj* 2012;345:e7570. doi: 10.1136/bmj.e7570 [published Online First: 2012/11/22]
18. Rathi VK, Strait KM, Gross CP, et al. Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials* 2014;15:384. doi: 10.1186/1745-6215-15-384 [published Online First: 2014/10/04]
19. Tannenbaum S, Ross JS, Krumholz HM, et al. Early Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators. *Annals of internal medicine* 2018;169(8):586-88. doi: 10.7326/m18-0723 [published Online First: 2018/07/18]
20. Endorsement of Data Sharing by Authors and High-Impact Medical Journals in China: A Survey of Authors and Assessment of Journal Online Instructions. The 8th International Congress on Peer Review and Biomedical Publication; 2017; Chicago.
21. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. The 6th International Congress on Peer Review and Biomedical Publication; 2009; Vancouver.
22. Griswold M, Laine C, Mulrow C, et al. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. The 7th International Congress on Peer Review and Biomedical Publication. Chicago, 2013.
23. Rowhani-Farid A, Barnett AG. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ open* 2016;6(10):e011784. doi: 10.1136/bmjopen-2016-011784 [published Online First: 2016/10/16]
24. Murugiah K, Ritchie JD, Desai NR, et al. Availability of Clinical Trial Data From Industry-Sponsored Cardiovascular Trials. *Journal of the American Heart Association* 2016;5(4):e003307. doi: 10.1161/jaha.116.003307 [published Online First: 2016/04/22]
25. Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *Journal of clinical epidemiology* 2019;109:111-16. doi: 10.1016/j.jclinepi.2019.01.012 [published Online First: 2019/02/10]
26. Kaufmann I, Gattrell WT. Assessment of journal compliance with data sharing guidelines from the International Committee of Medical Journal Editors (ICMJE) / in Original Abstracts from the 2019 European Meeting of ISMPP. *Current Medical Research and Opinion* 2019;35(sup2):31-41. doi: 10.1080/03007995.2019.1583496
27. Papageorgiou SN, Antonoglou GN, Martin C, et al. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. *Journal of orthodontics* 2019;46(2):101-09. doi: 10.1177/1465312519842315 [published Online First: 2019/05/09]
28. Kemper JM, Rolnik DL, Mol BWJ, et al. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. *Fertility and sterility* 2020;114(6):1322-29. doi: 10.1016/j.fertnstert.2020.05.020 [published Online First: 2020/08/11]
29. Johnson AL, Torgerson T, Skinner M, et al. An assessment of transparency and reproducibility-related research practices in otolaryngology. *The Laryngoscope* 2020;130(8):1894-901. doi: 10.1002/lary.28322 [published Online First: 2019/10/09]
30. Bergeris A, Tse T, Zarin DA. Trialists' Intent to Share Individual Participant Data as Disclosed at ClinicalTrials.gov. *Jama* 2018;319(4):406-08. doi: 10.1001/jama.2017.20581

31. Statham EE, White SA, Sonwane B, et al. Primed to comply: Individual participant data sharing statements on ClinicalTrials.gov. *PloS one* 2020;15(2):e0226143. doi: 10.1371/journal.pone.0226143 [published Online First: 2020/02/19]
32. Mayer CS, Williams N, Gold S, et al. Sharing of Individual Participant Data from Clinical Trials: General Comparison and HIV Use Case. *AMIA Annual Symposium proceedings AMIA Symposium* 2019;2019:647-54. [published Online First: 2020/04/21]
33. Ali J, Califf R, Sugarman J. Anticipated Ethics and Regulatory Challenges in PCORnet: The National Patient-Centered Clinical Research Network. *Accountability in research* 2016;23(2):79-96. doi: 10.1080/08989621.2015.1023951 [published Online First: 2015/07/21]
34. Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15(2):125-34. doi: 10.1111/1467-8519.00220 [published Online First: 2001/11/08]
35. Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *Journal of clinical epidemiology* 2019;106:60-69. doi: 10.1016/j.jclinepi.2018.10.006 [published Online First: 2018/10/22]
36. Godolphin PJ, Bath PM, Montgomery AA. Short email with attachment versus long email without attachment when contacting authors to request unpublished data for a systematic review: a nested randomised trial. *BMJ open* 2019;9(1):e025273. doi: 10.1136/bmjopen-2018-025273 [published Online First: 2019/02/02]
37. Howe N, Giles E, Newbury-Birch D, et al. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *Journal of health services research & policy* 2018;23(2):123-33. doi: 10.1177/1355819617751555 [published Online First: 2018/04/15]
38. Merson L, Phong TV, Nhan le NT, et al. Trust, Respect, and Reciprocity: Informing Culturally Appropriate Data-Sharing Practice in Vietnam. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):251-63. doi: 10.1177/1556264615592387 [published Online First: 2015/08/25]
39. Cheah PY, Tangseefa D, Somsaman A, et al. Perceived Benefits, Harms, and Views About How to Share Data Responsibly: A Qualitative Study of Experiences With and Attitudes Toward Data Sharing Among Research Staff and Community Representatives in Thailand. *Journal of empirical research on human research ethics : JERHRE* 2015;10(3):278-89. doi: 10.1177/1556264615592388 [published Online First: 2015/08/25]
40. Broes S, Verbaanderd C, Casteels M, et al. Sharing of Clinical Trial Data and Samples: The Cancer Patient Perspective. *Frontiers in medicine* 2020;7:33. doi: 10.3389/fmed.2020.00033 [published Online First: 2020/03/03]
41. Jones CW, Roberts BW, Platts-Mills TF. Patient Perspectives on Sharing Deidentified Trial Data. *Annals of internal medicine* 2016;165(10):748-49. doi: 10.7326/l16-0089 [published Online First: 2016/11/15]
42. Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *N Engl J Med* 2018;378(23):2202-11. doi: 10.1056/NEJMsa1713258 [published Online First: 2018/06/07]
43. Colombo C, Roberto A, Krleza-Jeric K, et al. Sharing individual participant data from clinical studies: a cross-sectional online survey among Italian patient and citizen groups. *BMJ open* 2019;9(2):e024863. doi: 10.1136/bmjopen-2018-024863 [published Online First: 2019/02/21]

44. Spence O, Onwuchekwa Uba R, Shin S, et al. Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials* 2018;19(1):269. doi: 10.1186/s13063-018-2651-2 [published Online First: 2018/05/05]
45. Medical Journal Editor Perspectives on Sharing Results Data From Patient-Oriented Research: A WAME Survey. 7th Peer Review Congress; 2009; Vancouver.
46. Promoting transparency in clinical research: Systematic review of disclosure and data-sharing policies in surgical journals. ASiT; 2014.
47. Data sharing trends and integrity in the era of open science: a qualitative analysis of journals in top-ranked therapeutic areas. 13th Annual Meeting of ISMPP; 2017.
48. Vidal-Infer A, Tarazona B, Alonso-Arroyo A, et al. Public availability of research data in dentistry journals indexed in Journal Citation Reports. *Clinical oral investigations* 2018;22(1):275-80. doi: 10.1007/s00784-017-2108-0 [published Online First: 2017/03/28]
49. Vassar M, Jellison S, Wendelbo H, et al. Data sharing practices in randomized trials of addiction interventions. *Addictive behaviors* 2020;102:106193. doi: 10.1016/j.addbeh.2019.106193 [published Online First: 2019/11/27]
50. Nutu D, Gentili C, Naudet F, et al. Open science practices in clinical psychology journals: An audit study. *Journal of abnormal psychology* 2019;128(6):510-16. doi: 10.1037/abn0000414 [published Online First: 2019/08/02]
51. Gorman DM. Use of publication procedures to improve research integrity by addiction journals. *Addiction (Abingdon, England)* 2019;114(8):1478-86. doi: 10.1111/add.14604 [published Online First: 2019/03/10]
52. Gorman DM, Ferdinand AO. High impact nutrition and dietetics journals' use of publication procedures to increase research transparency. *Research Integrity and Peer Review* 2020;5(1):12. doi: 10.1186/s41073-020-00098-9
53. Almaqrami BS, Hua F, Liu Y, et al. Research waste-related editorial policies of leading dental journals: Situation 2018. *Oral diseases* 2020;26(3):696-706. doi: 10.1111/odi.13257 [published Online First: 2019/12/18]
54. Siebert M, Gaba JF, Caquelin L, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. *BMJ open* 2020;10(5):e038887. doi: 10.1136/bmjopen-2020-038887 [published Online First: 2020/06/01]
55. Aleixandre-Benavent R, Sapena AF, Ferrer SC, et al. Policies regarding public availability of published research data in pediatrics journals. *Scientometrics* 2019;118(2):439-51. doi: 10.1007/s11192-018-2978-1
56. DeVito NJ, French L, Goldacre B. Noncommercial Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *Jama* 2018;319(16):1721-23. doi: 10.1001/jama.2018.2841 [published Online First: 2018/05/02]
57. Whitlock EP, Dunham KM, DiGioia K, et al. Noncommercial US Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *JAMA network open* 2019;2(1):e187498. doi: 10.1001/jamanetworkopen.2018.7498 [published Online First: 2019/01/27]
58. Rollando P, Parc C, Naudet F, et al. [Data sharing policies of clinical trials funders in France]. *Therapie* 2020;75(6):527-36. doi: 10.1016/j.therap.2020.04.001 [published Online First: 2020/05/25]
59. Goldacre B, Lane S, Mahtani KR, et al. Pharmaceutical companies' policies on access to trial data, results, and methods: audit study. *Bmj* 2017;358:j3334. doi: 10.1136/bmj.j3334 [published Online First: 2017/07/28]

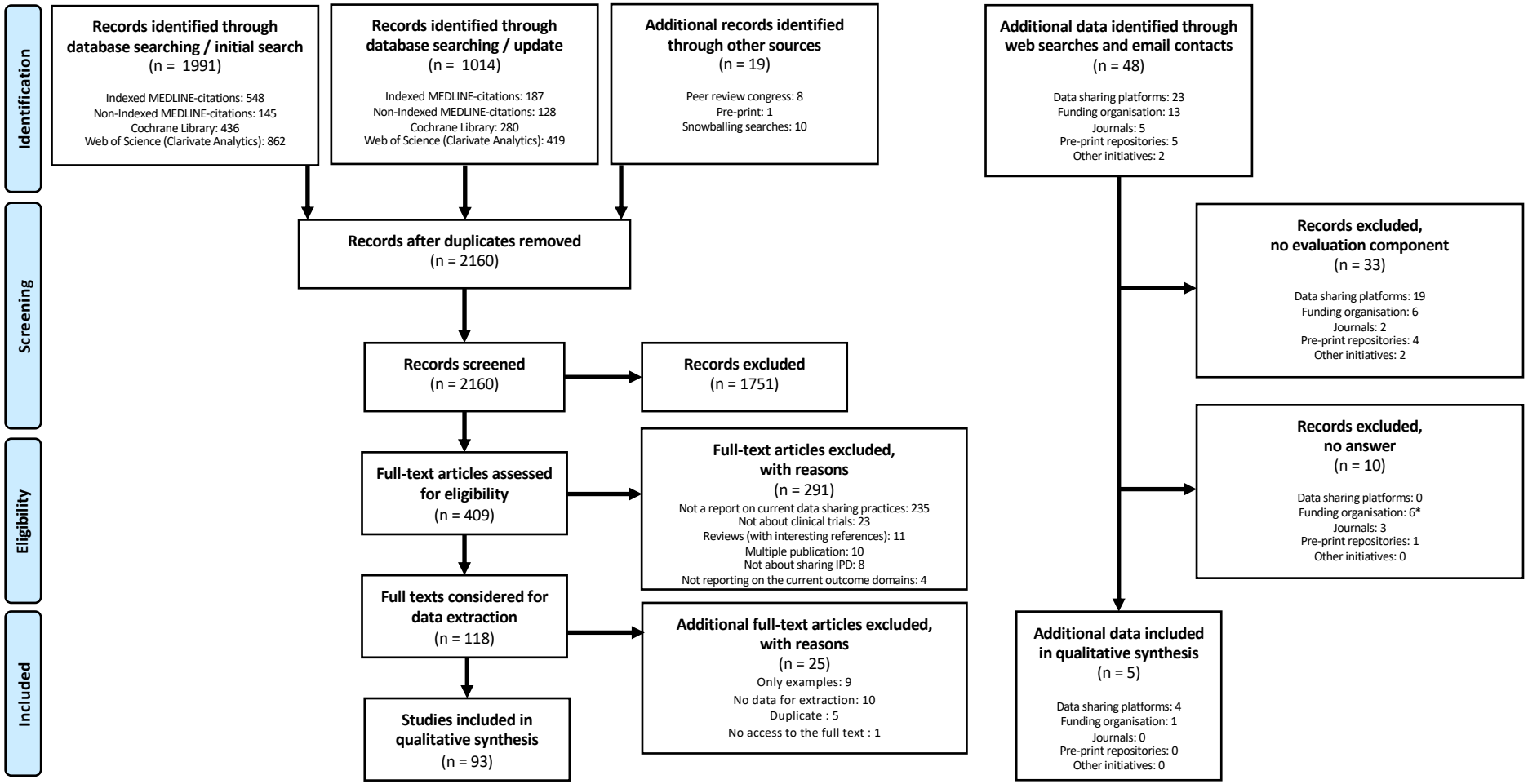
60. Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. *BMC medicine* 2018;16(1):165. doi: 10.1186/s12916-018-1154-z [published Online First: 2018/09/29]
61. Gaba JF, Siebert M, Dupuy A, et al. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PloS one* 2020;15(8):e0237464. doi: 10.1371/journal.pone.0237464 [published Online First: 2020/08/21]
62. Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005;352(21):2202-10. doi: 10.1056/NEJMsa044115 [published Online First: 2005/05/27]
63. Bosserdt M, Hamm B, Dewey M. Clinical trials in radiology and data sharing: results from a survey of the European Society of Radiology (ESR) research committee. *European radiology* 2019;29(9):4794-802. doi: 10.1007/s00330-019-06105-y [published Online First: 2019/02/28]
64. Savage CJ, Vickers AJ. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PloS one* 2009;4(9):e7078. doi: 10.1371/journal.pone.0007078
65. Naudet F, Sakarovitch C, Janiaud P, 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 The BMJ and PLOS Medicine. *Bmj* 2018;360:k400. doi: 10.1136/bmj.k400 [published Online First: 2018/02/15]
66. Rowhani-Farid A, Aldcroft A, Barnett AG. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Royal Society open science* 2020;7(3):191818. doi: 10.1098/rsos.191818 [published Online First: 2020/04/10]
67. Hee SW, Dritsaki M, Willis A, et al. Development of a repository of individual participant data from randomized controlled trials of therapists delivered interventions for low back pain. *European Journal of Pain* 2017;21(5):815-26. doi: https://doi.org/10.1002/ejp.984
68. Nevitt SJ, Marson AG, Davie B, et al. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *Bmj* 2017;357:j1390. doi: 10.1136/bmj.j1390 [published Online First: 2017/04/07]
69. Azar M, Benedetti A, Riehm KE, et al. Individual participant data meta-analyses (IPDMA): data contribution was associated with trial corresponding author country, publication year, and journal impact factor. *Journal of clinical epidemiology* 2020;124:16-23. doi: https://doi.org/10.1016/j.jclinepi.2020.03.026
70. Mayo-Wilson E, Doshi P, Dickersin K. Are manufacturers sharing data as promised? *BMJ : British Medical Journal* 2015;351:h4169. doi: 10.1136/bmj.h4169
71. Kawahara T, Fukuda M, Oba K, et al. Meta-analysis of randomized clinical trials in the era of individual patient data sharing. *International journal of clinical oncology* 2018;23(3):403-09. doi: 10.1007/s10147-018-1237-z
72. Villain B, Dechartres A, Boyer P, et al. Feasibility of individual patient data meta-analyses in orthopaedic surgery. *BMC medicine* 2015;13(1):131. doi: 10.1186/s12916-015-0376-6
73. Veroniki AA, Ashoor HM, Le SPC, et al. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *Journal of clinical epidemiology* 2019;113:176-88. doi: 10.1016/j.jclinepi.2019.05.031 [published Online First: 2019/06/04]
74. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *Bmj* 2012;344:d7762. doi: 10.1136/bmj.d7762 [published Online First: 2012/01/05]

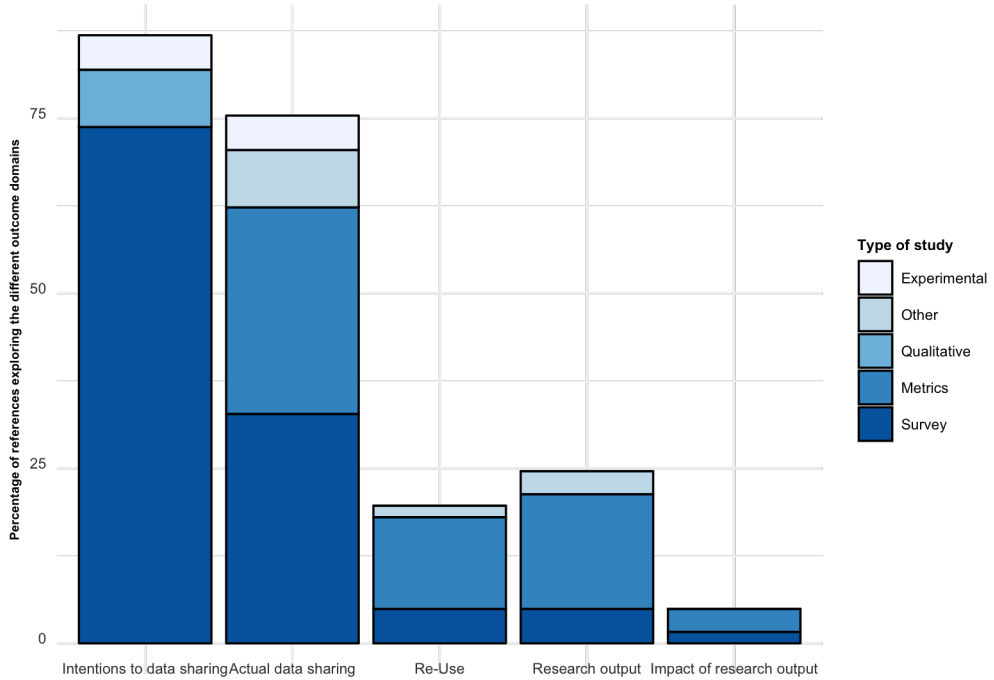
75. Geifman N, Bollyky J, Bhattacharya S, et al. Opening clinical trial data: are the voluntary data-sharing portals enough? *BMC medicine* 2015;13:280. doi: 10.1186/s12916-015-0525-y [published Online First: 2015/11/13]
76. Sydes MR, Johnson AL, Meredith SK, et al. Sharing data from clinical trials: the rationale for a controlled access approach. *Trials* 2015;16(1):104. doi: 10.1186/s13063-015-0604-6
77. Ross JS, Ritchie JD, Finn E, et al. Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ open* 2016;6(9):e012769. doi: 10.1136/bmjopen-2016-012769 [published Online First: 2016/09/28]
78. Kochhar S, Knoppers B, Gamble C, et al. Clinical trial data sharing: here's the challenge. *BMJ open* 2019;9(8):e032334. doi: 10.1136/bmjopen-2019-032334 [published Online First: 2019/08/24]
79. So D, Knoppers BM. Ethics approval in applications for open-access clinical trial data: An analysis of researcher statements to clinicalstudydatarequest.com. *PloS one* 2017;12(9):e0184491. doi: 10.1371/journal.pone.0184491 [published Online First: 2017/09/09]
80. Vaduganathan M, Nagarur A, Qamar A, et al. Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation* 2018;137(9):938-47. doi: 10.1161/circulationaha.117.031883 [published Online First: 2017/11/15]
81. Strom BL, Buyse M, Hughes J, et al. Data sharing, year 1--access to data from industry-sponsored clinical trials. *N Engl J Med* 2014;371(22):2052-4. doi: 10.1056/NEJMp1411794 [published Online First: 2014/10/16]
82. Schmidt H, Caldwell B, Mollet A, et al. An Industry Experience with Data Sharing. *N Engl J Med* 2018;379(21):2081-82. doi: 10.1056/NEJMc1805610 [published Online First: 2018/11/22]
83. Strom BL, Buyse ME, Hughes J, et al. Data Sharing - Is the Juice Worth the Squeeze? *N Engl J Med* 2016;375(17):1608-09. doi: 10.1056/NEJMp1610336 [published Online First: 2016/10/27]
84. Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) Project — A Mechanism for Data Sharing. *New England Journal of Medicine* 2016;375(5):403-05. doi: 10.1056/NEJMp1607342
85. Ross JS, Ritchie JD, Bamford S, et al. Sharing Data Through the Yale University Open Data Access (YODA) Project: Early Experience. The 8th International Congress on Peer Review and Biomedical Publication. Chicago, 2017.
86. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Scientific data* 2018;5:180268. doi: 10.1038/sdata.2018.268 [published Online First: 2018/11/28]
87. Navar AM, Pencina MJ, Rymer JA, et al. Use of Open Access Platforms for Clinical Trial Data. *Jama* 2016;315(12):1283-84. doi: 10.1001/jama.2016.2374
88. Boutron I, Dechartres A, Baron G, et al. Sharing of Data From Industry-Funded Registered Clinical Trials. *Jama* 2016;315(24):2729-30. doi: 10.1001/jama.2016.6310 [published Online First: 2016/07/02]
89. Miller J, Ross JS, Wilenzick M, et al. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *Bmj* 2019;366:l4217. doi: 10.1136/bmj.l4217 [published Online First: 2019/07/12]
90. Coady SA, Mensah GA, Wagner EL, et al. Use of the National Heart, Lung, and Blood Institute Data Repository. *N Engl J Med* 2017;376(19):1849-58. doi: 10.1056/NEJMsa1603542 [published Online First: 2017/04/13]

91. Shmueli-Blumberg D, Hu L, Allen C, et al. The national drug abuse treatment clinical trials network data share project: website design, usage, challenges, and future directions. *Clinical trials (London, England)* 2013;10(6):977-86. doi: 10.1177/1740774513503522 [published Online First: 2013/10/03]
92. Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. *Clinical trials (London, England)* 2018;15(4):413-23. doi: 10.1177/1740774518769655 [published Online First: 2018/04/21]
93. Bertagnolli MM, Sartor O, Chabner BA, et al. Advantages of a Truly Open-Access Data-Sharing Model. *N Engl J Med* 2017;376(12):1178-81. doi: 10.1056/NEJMsb1702054 [published Online First: 2017/03/23]
94. Asare AL, Carey VJ, Rotrosen D, et al. Clinical trial data access: Opening doors with TrialShare. *The Journal of allergy and clinical immunology* 2016;138(3):724-26. doi: 10.1016/j.jaci.2016.05.034 [published Online First: 2016/07/31]
95. MRC. MRC Clinical Trials Transparency Review 2017 [Available from: https://mrc.ukri.org/documents/pdf/clinical-trials-review-2017/ accessed 22/11/2019.
96. Zhu CS, Pinsky PF, Moler JE, et al. Data sharing in clinical trials: An experience with two large cancer screening trials. *PLoS medicine* 2017;14(5):e1002304. doi: 10.1371/journal.pmed.1002304 [published Online First: 2017/05/26]
97. Chalmers I, Altman DG, McHaffie H, et al. Data sharing among data monitoring committees and responsibilities to patients and science. *Trials* 2013;14:102. doi: 10.1186/1745-6215-14-102 [published Online First: 2013/06/21]
98. Tudur Smith C, Nevitt S, Appelbe D, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials* 2017;18(1):319. doi: 10.1186/s13063-017-2067-4 [published Online First: 2017/07/18]
99. Lovato L. Exploring the potential of data sharing: The SPRINT Challenge in Abstracts from the SCT 39th Annual Meeting (2018). *Clinical trials (London, England)* 2018;15(2_suppl):S35-s192. doi: 10.1177/1740774518790846 [published Online First: 2018/08/22]
100. Ebrahim S, Sohani ZN, Montoya L, et al. Reanalyses of randomized clinical trial data. *Jama* 2014;312(10):1024-32. doi: 10.1001/jama.2014.9646 [published Online First: 2014/09/10]
101. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PloS one* 2007;2(3):e308. doi: 10.1371/journal.pone.0000308 [published Online First: 2007/03/22]
102. ECRIN. ECRIN metadata repository [Available from: https://ecrin.org/tools/clinical-research-metadata-repository accessed 01.07.2021 2021.
103. Doshi P, Shamseer L, Jones M, et al. Restoring biomedical literature with RIAT. *Bmj* 2018;361:k1742. doi: 10.1136/bmj.k1742 [published Online First: 2018/04/28]
104. Takata J, Sondo P, Humphreys GS, et al. The WorldWide Antimalarial Resistance Network Clinical Trials Publication Library: A Live, Open-Access Database of Plasmodium Treatment Efficacy Trials. *The American journal of tropical medicine and hygiene* 2020;103(1):359-68. doi: 10.4269/ajtmh.19-0706 [published Online First: 2020/05/21]
105. Pisani E, Botchway S. Sharing individual patient and parasite-level data through the WorldWide Antimalarial Resistance Network platform: A qualitative case study. *Wellcome Open Res* 2017;2:63-63. doi: 10.12688/wellcomeopenres.12259.1
106. Crossmark [Available from: https://www.crossref.org/services/crossmark/ accessed 01.11.2021 2021.

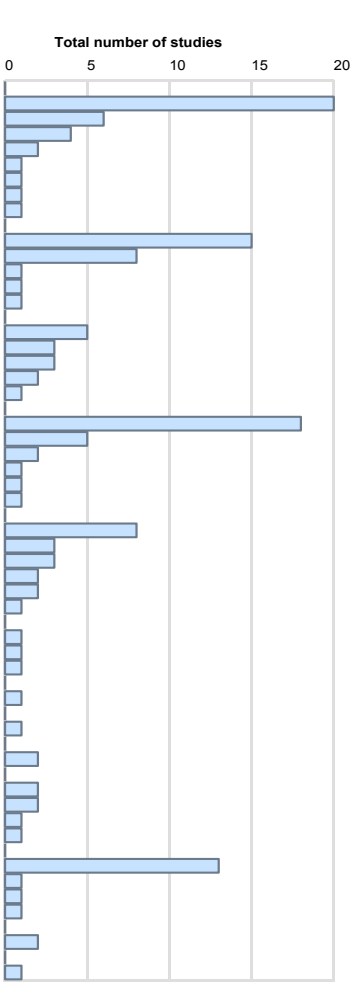
107. Devereaux PJ, Guyatt G, Gerstein H, et al. Toward Fairness in Data Sharing. *N Engl J Med* 2016;375(5):405-7. doi: 10.1056/NEJMp1605654 [published Online First: 2016/08/16]

108. Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic reviews* 2012;1:2. doi: 10.1186/2046-4053-1-2 [published Online First: 2012/05/17]

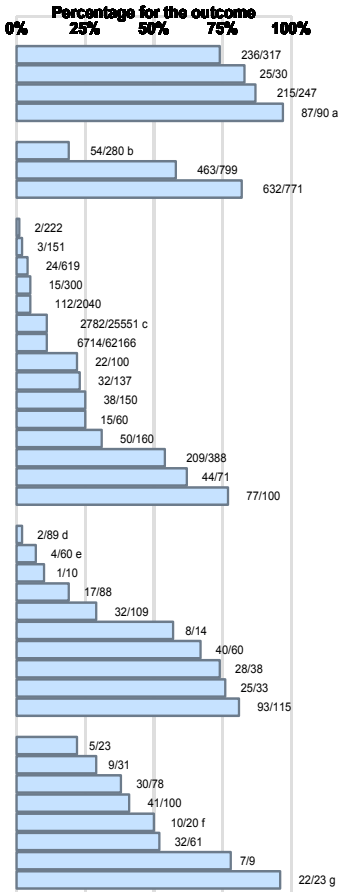

Identification
Screening
Eligibility
Included
Records identified through database searching / initial search (n = 1991)
Indexed MEDLINE-citations: 548
Non-Indexed MEDLINE-citations: 145
Cochrane Library: 436
Web of Science (Clarivate Analytics): 862
Records identified through database searching / update (n = 1014)
Indexed MEDLINE-citations: 187
Non-Indexed MEDLINE-citations: 128
Cochrane Library: 280
Web of Science (Clarivate Analytics): 419
Additional records identified through other sources (n = 19)
Peer review congress: 8
Pre-print: 1
Snowballing searches: 10
Additional data identified through web searches and email contacts (n = 48)
Data sharing platforms: 23
Funding organisation: 13
Journals: 5
Pre-print repositories: 5
Other initiatives: 2
Records after duplicates removed (n = 2160)
Records screened (n = 2160)
Records excluded (n = 1751)
Records excluded, no evaluation component (n = 33)
Data sharing platforms: 19
Funding organisation: 6
Journals: 2
Pre-print repositories: 4
Other initiatives: 2
Full-text articles assessed for eligibility (n = 409)
Full-text articles excluded, with reasons (n = 291)
Not a report on current data sharing practices: 235
Not about clinical trials: 23
Reviews (with interesting references): 11
Multiple publication: 10
Not about sharing IPD: 8
Not reporting on the current outcome domains: 4
Records excluded, no answer (n = 10)
Data sharing platforms: 0
Funding organisation: 6*
Journals: 3
Pre-print repositories: 1
Other initiatives: 0
Full texts considered for data extraction (n = 118)
Additional full-text articles excluded, with reasons (n = 25)
Only examples: 9
No data for extraction: 10
Duplicate : 5
No access to the full text : 1
Additional data included in qualitative synthesis (n = 5)
Data sharing platforms: 4
Funding organisation: 1
Journals: 0
Pre-print repositories: 0
Other initiatives: 0
Studies included in qualitative synthesis (n = 93)


Percentage of references exploring the different outcome domains
75
50
25
0
Intentions to data sharing
Actual data sharing
Re-Use
Research output
Impact of research output
Type of study
Experimental
Other
Qualitative
Metrics
Survey

| Outcome domains | Outcomes | Experimental | Survey | Metrics | Qualitative | Other |
|---|---|---|---|---|---|---|
| Intentions to data sharing | **For trialists** | . | . | . | . | . |
| | Intentions to share data | 2 | 18 | . | . | . |
| | Existence of a data sharing plan/data sharing statement | . | 6 | . | . | . |
| | Type of data sharing plan | . | 4 | . | . | . |
| | Support for data sharing | . | 2 | . | . | . |
| | Willingness to store IPD in a central repository, conditions for storage | . | 1 | . | . | . |
| | Most pressing ethical issues | . | 1 | . | . | . |
| | Response to a request for data sharing and time to response | 1 | . | . | . | . |
| | Preparedness for DS | 1 | . | . | . | . |
| | **For publishers/funders** | . | . | . | . | . |
| | Existence of a data sharing policy/intent to share | . | 15 | . | . | . |
| | Type of data sharing policy | . | 8 | . | . | . |
| | Journals requiring participant level data | . | 1 | . | . | . |
| | Prohibiting data sharing in agreements with industry | . | 1 | . | . | . |
| | Existence of a data sharing policy for clinical trial units | . | 1 | . | . | . |
| | **For trial participants** | . | . | . | . | . |
| | Willingness to share data | . | 4 | . | 1 | . |
| | Barriers to share data | . | 2 | . | 1 | . |
| | Patient Opinions on the Release of Deidentified Individual-Patient Data | . | 3 | . | . | . |
| | Views, experience and attitudes towards DS | . | . | . | 2 | . |
| | Information in consent forms | . | 1 | . | . | . |
| Actual data sharing | **For re-users (e.g. IPD meta-analysts)** | . | . | . | . | . |
| | Data availability (yes/no) | 3 | 13 | . | . | 2 |
| | Data release after a request | . | 4 | 1 | . | . |
| | Data availability (time) | 1 | . | 1 | . | . |
| | Data availability (completeness) | 1 | . | . | . | . |
| | Rate of IPD meta-analyses / all requests | . | . | 1 | . | . |
| | Reasons for request, research plan and experience of DS | . | 1 | . | . | . |
| | **For repositories/platforms** | . | . | . | . | . |
| | Data release after a request | . | 2 | 5 | . | 1 |
| | Approval of data release | . | 2 | 1 | . | . |
| | Data availability | . | 3 | . | . | . |
| | Number of downloads | . | . | 2 | . | . |
| | Number of DS agreement, speed of data availability | . | 1 | 1 | . | . |
| | Database access | . | . | 1 | . | . |
| | **For other stakeholders** | . | . | . | . | . |
| | Number of downloads and page hits for any public access | . | . | 1 | . | . |
| | Re-use of data by independent data monitoring committees | . | . | . | . | 1 |
| | Cost of preparing the data for data generators | . | . | . | . | 1 |
| Re-use | **Further additional analyses** | . | . | . | . | . |
| | Approved proposals by subject areas | . | . | 1 | . | . |
| | **IPD meta-analyses** | . | . | . | . | . |
| | Proposals for IPD-meta-analyses and one concrete example | . | . | 1 | . | . |
| | **Re-analyses** | . | . | . | . | . |
| | Reproducibility on primary outcomes | . | 2 | . | . | . |
| | **Any type of re-use** | . | . | . | . | . |
| | Type of re-use | . | . | 2 | . | . |
| | Progression of the analysis | . | . | 1 | . | 1 |
| | Listing of type of studies performed | . | . | 1 | . | . |
| | Project completion | . | 1 | . | . | . |
| Output from data sharing | **Published re-use** | . | . | . | . | . |
| | Publication of re-use (papers) | . | 2 | 9 | . | 2 |
| | Manuscripts in peer review | . | . | 1 | . | . |
| | Request discontinued | . | . | 1 | . | . |
| | Communication of re-use (oral presentation, posters) | . | . | . | . | 1 |
| Impact of research output | **Quantitative metrics for published articles** | . | . | . | . | . |
| | Citation metrics | . | 1 | 1 | . | . |
| | **Identification of a new finding** | . | . | . | . | . |
| | Crowd voting | . | 1 | . | . | . |



| Reference | Type | Population | Time point |
|---|---|---|---|
| **A: Intent to share: surveys of trialists** | | | |
| Rathi, 2012 | Overall | Trials published in 6 high impact journals | 2010-2011 |
| Tudur-Smith, 2014 | Overall (central storage of their IPD) | Reviewers of Cochrane IPD meta-analysis group | 2011 |
| Yuanyuan, 2017 | Overall (endorsment of DS) | Trials publihed in Chinese Medical Journal | 2016 |
| Tannenbaum, 2018 | For an IPD meta-analysis | Trials published in 3 high impact journals with data sharing policies | 2012-2016 |
| **B: Intent to share: surveys of trial participants** | | | |
| Colombo, 2017 | Having knowlegde about and being in favor of data sharing | Italian patients and citizen groups | 2017 |
| Jones, 2016 | Favor or strongly favor data sharing | Patients in a Usemergency department | 2015 |
| Mello, 2018 | Perception that the benefits of data sharing outweighed the negative aspects | Patients from 3 US medical centers | Unclear |
| **C: Intent to share: data sharing statements** | | | |
| Kemper, 2020 | Overall | Reproductive endocrinology and infertility articles (study mix) | 2013, 2018 |
| Johnson, 2020 | Overall | 300 otolaryngology research studies | 2014-2018 |
| Gabelica, 2019 | Overall | RCTs in 7 high-ranked anesthesiology journals | 2014-2016 |
| Papageorgios, 2019 | Open data | Trials in orthodontics and periodontics | 2017-2018 |
| Statham, 2020 | Overall | CT.gov | 2018 |
| Bergeris, 2018 | Overall | CT.gov | Up to august 2017 |
| Mayer, 2019 | Studies with a data sharing plan | CT.gov | 2015-2018 |
| Siebert, 2020b | Overall | ICMJE affiliates (after policy) | 2019 |
| Kaufman, 2019 | Overall | RCTs in 11 selected journals (before policy) | 2018 |
| Kaufman, 2019b | Overall | RCTs in 11 selected journals (after policy) | 2018 |
| Murugiah, 2016 | Data made available | Clinical trials (> 5000 patients), from clinicaltrials.gov | Up to 2015 |
| Rowhani-Farid, 2016 | Overall | The BMJ (all research papers) | 2009-2015 |
| Griswold M, 2013 | Overall | Ann. Int. Med. (all research papers) | 2008-2012 |
| Laine, 2009 | Overall | Ann. Int. Med. (all research papers) | 2008 |
| Siebert, 2020 | Overall | ICMJE members (after policy) | 2019 |
| **D: Intent to share: journal data sharing policies** | | | |
| Krleza-Jeric, 2009 | Require IPD | Members of World Association of Medical Editors | 2009 |
| Chickramane, 2017 | Require IPD | 15 oncology, 15 central nervous system, 15 cardiology/endocrinology and 15 respiratory journals | Unknown |
| Chapman, 2014 | Data sharing policy | High impact surgical journals | 2009-2012 |
| Vidal-Infer, 2018 | Accept IPD as a complementary material | Dental journals | 2014 |
| Almaqrami, 2020 | Data sharing policy | Dental journals | 2018 |
| Vasar, 2020 | Data sharing policy | 15 high-impact addiction journals | 2013-2018 |
| Nutu, 2019 | Data sharing policy | Clinical psychology journals | 2017 |
| Gorman, 2019 | Data sharing policy | High impact addiction journals | 2018 |
| Gorman, 2020 | Data sharing policy | Nutrition and dietetics journals | 2018 |
| Alexandre-Benavent, 2019 | Data sharing policy | Paediatric journals | 2012-2016 |
| **E: Intent to share: funders and clinical trial units** | | | |
| Hopkins, 2016 | Data sharing policy | UK CTUs | 2014 |
| Rollando, 2020 | Data sharing policy | Clinical trial funders in France | 2019 |
| Gaba, 2020 | Data sharing policy | Non-commercial | 2016-2018 |
| Gaba, 2020b | Data sharing policy | Commercial funders | 2016-2018 |
| de Vito, 2018 | Data sharing requirement | Top non commercial funders | 2017 |
| Hopkins, 2018 | Data sharing policy | Clinical trials sponsored by the pharmaceutical industry published in the top 10 medical journals | 2015 |
| Whitlock, 2019 | Data sharing policy | Non-commercial funders in the US | 2018 |
| Goldacre, 2017 | Data sharing policy | Top 25 pharmaceutical companies by revenue | 2016 |



**A: Actual data sharing by re-users: surveys related to published studies**

| Reference | Data source | Sample selection | Time period |
|---|---|---|---|
| Vassar, 2020 | 15 high ranked addiction journals | Consecutive | 2013-2018 |
| Gabelica, 2019 | 7 high ranked addiction journals | Consecutive | 2014-2016 |
| Savage, 2009 | PLOS Medicine, PLOS clinical trials | Unclear | 2009 |
| Yuanyuan, 2017 | Chinese Medical Journal | Consecutive | 2016 |
| Hopkins, 2018 | Top 10 general and internal medical journals | Consecutive, industry-sponsored trials | 2015 |
| Rowhani-Farid, 2016 | BMJ | Random, subsample RCTs | 2009-2015 |
| Rathi, 2012 | 6 general medical journals | Consecutive | 2010-2011 |
| Naudet, 2018 | BMJ, PLOS Medicine | Consecutive, RCTs | 2013-2016 |
| Tannenbaum, 2018 | BMJ, PLOS Medicine, Ann. Inn. Med. | Selected, partly restricted to authors willing to share data | 2012-2016 |

**B: Actual data sharing by re-users: data related to IPD meta-analyses**

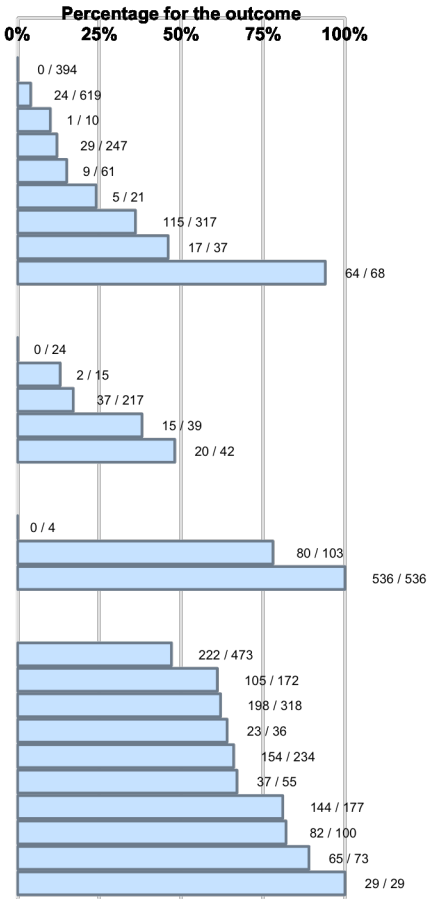
| Reference | Number of studies with IPD sharing | Time point | Comment |
|---|---|---|---|
| Mayo-Wilson, 2015 | 0/24 (0%) | Contacted in 2014 | Commercial sponsor, trials with one medicinal product |
| Kawahara, 2018 | 2/15 (13%) through CSDR | Unknown | 13 requested from authors directly, project in progress |
| Villain, 2015 | 37/217 (17%) | RCTs with results published since 2000 | |
| Nevitt,2017 | 15/35 (38%), 4/15 through CSDR | End 2015 | |
| Hee, 2016 | 20/42 (48%) | RCTs until 2011 | |

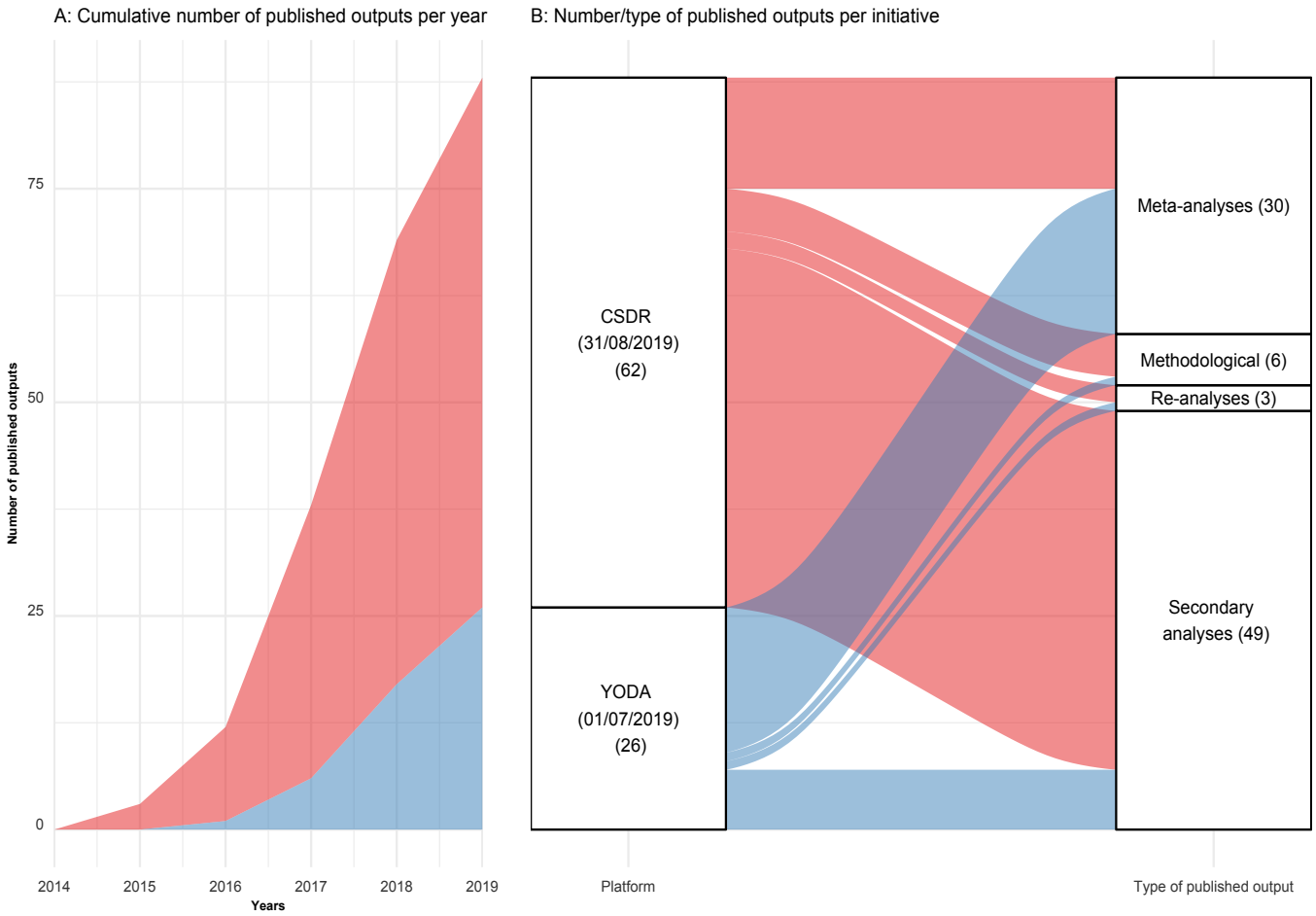
**C: Actual data sharing by re-users: repositories/platforms from the viewpoint of the user**

| Reference | Repository_platform | Time point of submission | Comment |
|---|---|---|---|
| Geifman, 2015 | CSDR | 12/2014-1/2015 | |
| Sydes, 2015 | MRC CTU | 2012-2014 | 103 requests to 54 trials |
| Ross, 2016 | BioLINCC | 2007-2014 | Survey of investigators who received access to BioLINCC |

**D: Actual data sharing by re-users: survey of repositories/platforms**

| Reference | Platform | Time point of assessment | Comment |
|---|---|---|---|
| Kochar, 2019 | CSDR | 2014-2019 | |
| So, 2017 | CSDR | February 2017 | |
| Vaduganathan, 2018 | CSDR | May 2017 | |
| Strom, 2014 | CSDR | May 2014 | |
| Navar, 2016 | CSDR, YODA, SOAR | December 2015 | |
| Schmidt, 2018 | CSDR | 2014-2017 | Boehringer-Ingelheim's studies |
| Strom, 2016 | CSDR | November 2015 | 177 for 237 trials |
| Ross, 2018 | YODA | August 2018 | |
| Ross, 2017 | YODA | June 2017 | 73 for 159 trials |
| Krumholz, 2016 | YODA | 2015 | |




A: Cumulative number of published outputs per year
B: Number/type of published outputs per initiative
Number of published outputs
0
25
50
75
2014
2015
2016
2017
2018
2019
Years
CSDR
(31/08/2019)
(62)
YODA
(01/07/2019)
(26)
Platform
Meta-analyses (30)
Methodological (6)
Re-analyses (3)
Secondary
analyses (49)
Type of published output

# Supplementary material 1: Information sources

For commercial sponsors, we considered:
1/ Clinical Study Data Request (CSDR),
2/ the Yale University Open Data Access Project (YODA),
3/ the Supporting Open Access for Researchers initiative (SOAR),
4/ ViVli.

For non-commercial sponsor, we considered:
1/ the National Institute of Mental Health, National Database for Clinical Trials Related to Mental Illness (NIMH NDCT),
2/ The National Institute of Health, Biologic Specimen and Data Repositories Information Coordinating Center (NIH BioLINCC),
3/ B2Share,
4/ Dryad,
5/ the Data Repository for University of Minnesota (Drum),
6/ EASY,
7/ Edinburgh DataShare,
8/ FigShare,
9/ the Inter-university Consortium for Political and Social Research (ICPSR),
10/ the Swedish National data Service (SND),
11/ the University Medical Hospital Information Network (UMIN),
12/ Zenodo,
13/ the Early Breast Cancer Trialists' Collaborative Group (EBCTG),
14/ FreeBird,
15/ Traumatic Brain Injury – International Mission for Prognosis and Analysis of Clinical trials in TBI (TBI-IMPACT),
16/ Melanoma Molecular Map Project (MMMP),
17/ National institute of Diabetes and Digestive and Kidney Diseases (NIDDK),
18/ Immune Tolerance Network TrialShare (ITN Trialshare),
19/ Child Abuse,
20/ Pooled Resource Open Access Clinical trials database (ProAct).

For the different funders:
1/ National Institute of Health (NIH US),
2/ European Commission (EC Europe),
3/ Medical Research Council (MRC UK),
3/ Le programme hospitalier de recherche clinique (DGOS France),
4/ L'Agence nationale de la recherche (ANR France),
5/ Department of Defense (US DoD),
6/ Wellcome Trust UK,

7/ Canadian Institutes of Health Research (CIHR Canada),
8/ National Health and Medical Research Council (NHMRC Australia),
9/ Deutsche Forschungsgemeinschaft (DFG Germany),
10/ National Natural Science Foundation of China (NSFC China),
11/ National Institute of Health Research (NIHR UK),
12/ Gates Foundation US.

# Supplementary material 2: Literature searches

The initial algorithm for the literature search, detailed in the registered protocol (osf.io/pb8cj), was updated on October 29th 2018 to include broader search terms.

| **Name of Database** | **Host, search interface** | **Initial search** Search date 2018-10-29 | | **Update search** Search date 2020-09-11 Publication year 2018-2020 | |
|---|---|---|---|---|---|
| | | **Update status of the database** | **Results** | **Update status of the database** | **Results** Publication year From 2018-2020 |
| Medline | Wolters Kluwer / Ovid | 1946 to October Week 3 2018 | 548 | 1946 to September Week 1 2020 | 187 |
| Medline daily update | | October 25, 2018 | | September 09, 2020 | |
| MEDLINE In-Process & Other Non-Indexed Citations | | | 145 | 1946 to September 09, 2020 | 128 |
| MEDLINE Epub Ahead of Print | | | | September 09, 2020 | |
| Cochrane Library: Cochrane Reviews | Wiley | Issue 10 of 12, October 2018 | 19 | Issue 9 of 12, September 2020 | 12 |
| Cochrane Protocols | | | 1 | | 0 |
| Cochrane Central Register of Controlled Trials | | Issue 9 of 12, September 2018 | 416 | Issue 9 of 12, September 2020 | 268 |
| Science Citation Index | Clarivate Analytics / Web of Science | 1945 –present (2018-10-26) | 862 | 2018 –present (2020-09-10) | 419 |
| Social Science Citation Index | | 1956-present (2018-10-26) | | | |

| | | | | |
|---|---|---|---|---|
| **Total with duplicates** | | **1991** | | **1014** |
| **Total without duplicates** | | **1544** | | **763** |
| **New citations 2018-2020 without overlap from initial search** | | | | **597** |
| **Total without overlap initial search and update search (see PRISMA flow diagram)** | | **2141 (1544 + 597)** | | |

**MEDLINE Databases: Host: Wolters Kluwer, search interface: Ovid**

**1. Indexed MEDLINE-citations:**
Search Strategy:

| # | **Searches** | **Results Initial search** Search date: 2018-10-29: MEDLINE 1946 to October Week 3 2018, MEDLINE Daily Update October 25, 2018 | **Results Update search** Search date: 2020-09-11: MEDLINE 1946 to September Week 1 2020, MEDLINE (Daily Update September 09, 2020. | **Annotations** |
|---|---|---|---|---|
| 1 | exp Access to Information/ | 6845 | 7597 | Concept data sharing: MeSH terms |
| 2 | Information Dissemination/ | 14697 | 16894 | |
| 3 | exp *"Information Storage and Retrieval"/ | 52187 | 58658 | |
| 4 | data collection/ | 87165 | 89553 | |
| 5 | datasets as topic/ | 2259 | 4417 | |
| 6 | or/1-5 | 157717 | 170739 | |
| 7 | exp clinical trial/ | 809623 | 868410 | Concept clinical trials: |

| | | | | |
|---|---|---|---|---|
| 8 | exp clinical trial as topic/ | 318580 | 345552 | MeSH terms or textwords |
| 9 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 879338 | 997736 | |
| 10 | Meta-Analysis as Topic/ | 16485 | 18279 | Concept Meta-analysis: MeSH |
| 11 | meta-analysis/ | 93492 | 119228 | |
| 12 | or/7-11 | 1489253 | 1650537 | Concept Clinical Trials OR Meta-analysis |
| 13 | 6 and 12 | 10206 | 11264 | Combination of concepts: data sharing (MeSH only) AND (clinical trials OR meta-analysis) |
| 14 | (data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 7822 | 9921 | Concept data sharing: Textwords |
| 15 | 13 and 14 | 325 | 422 | data sharing (MeSH terms) AND data sharing (textwords) AND (clinical trials OR meta-analysis):<br>1. interim result |
| 16 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 756 | 993 | Concept sharing IPD (textwords) |
| 17 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 23 | 38 | |
| 18 | or/16-17 | 772 | 1017 | |
| 19 | 12 and 18 | 129 | 179 | (Clinical trials OR meta-analysis) AND textwords for sharing IPD:<br>2. interim result |
| 20 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3196 | 4262 | Concept data sharing textwords |

| | | | | |
|---|---|---|---|---|
| 21 | 12 and 20 | 393 | 539 | (Clinical trials OR meta-analysis) AND textwords data sharing:<br>3. interim result |
| 22 | 15 or 19 or 21 | 557 | 738 | OR-combination of interim results |
| 23 | exp animals/ not humans/ | 4508403 | 4732433 | Exclusion of animals only |
| 24 | 22 not 23 | 552 | 732 | |
| 25 | limit 24 to (english or french or german or spanish) | 548 | 725 | Restriction to English, German, French, Spanish:<br>Final result for indexed Medline citations |
| | | | 187 | Update search:<br>limit 25 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Exp **term*/ = MeSH as major topic incl. narrower terms as major topic
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters
adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**2. Non-Indexed MEDLINE-citations:**

| # | **Searches** | **Results Initial search**<br>Search date: 2018-10-29:<br>MEDLINE In-Process & Other Non-Indexed | **Results Update search**<br>Search date: 2020-09-11:<br>MEDLINE Epub Ahead of Print September 09, 2020, MEDLINE | **Annotations** |
|---|---|---|---|---|

| | | Citations October 25, 2018, MEDLINE Epub Ahead of Print October 25, 2018 | In-Process & Other Non-Indexed Citations 1946 to September 09, 2020 | |
|---|---|---|---|---|
| 1 | ((individual* or patient* or participant*) adj6 data adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 215 | 323 | Concept data sharing |
| 2 | (IPD adj6 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 3 | 7 | |
| 3 | (data adj1 (share* or sharing* or reuse* or re-use* or reusing or re-using)).ti,ab,kf. | 1122 | 1564 | |
| 4 | or/1-3 | 1230 | 1727 | |
| 5 | exp clinical trial/ | 401 | 521 | Concept clinical trials |
| 6 | (randomi#ed or randomly or randomi#ation or ((random* or clinical) adj3 trial*)).ti,ab,kf. | 138560 | 174420 | |
| 7 | meta-analysis as topic/ | 1 | 0 | Concept meta-analysis |
| 8 | meta-analysis/ | 34 | 99 | |
| 9 | (meta-analy* or metaanaly*).ti,ab,kf. | 27362 | 37834 | |
| 10 | or/5-9 | 156727 | 199473 | Concept clinical trials OR meta-analysis |
| 11 | 4 and 10 | 146 | 191 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| 12 | limit 11 to (english or french or german or spanish) | 145 | 191 | Restriction to English, French, German, Spanish:<br>Final result for non-indexed Medline citations |
| | | | 128 | Update search:<br>limit 12 to yr="2018 - 2020" |

*Term*/ = MeSH (Medical subject heading
Exp *term*/ = exploded Mesh (incl. narrower terms)
Wildcards, Truncation:
# = replaces exact one character
* = zero or any number of characters

adj*n* = terms within *n* words in any order
ti,ab,kf = textword search in title, abstract, keyword heading word (author kewords)

**Cochrane Library** (Wiley):
- Cochrane Database of Systematic Reviews
- Cochrane Protocols
- Cochrane Central Register of Controlled Trials

| ID | Search | Annotations |
|---|---|---|
| #1 | ((data near share*) or (data near sharing*)):ti,ab,kw | Concept data sharing: Textword search in title, abstract, keywords |
| #2 | (data next share*) or (data next sharing*) | Concept data sharing: Textword search in fulltext |
| #3 | #1 or #2 | Concept data sharing. 1. Interim result |
| #4 | ((patient* or participant*) near individual*):ti,ab,kw | Concept Individual patient data sharing. 2. Interim result |
| #5 | data:ti,ab,kw | |
| #6 | (share* or sharing*):ti,ab,kw | |
| #7 | #4 and #5 and #6 | |
| #8 | (IPD near (share* or sharing*)):ti,ab,kw | Concept IPD sharing: 3. Interim result |
| #9 | #3 or #7 or #8 in Cochrane Reviews, Cochrane Protocols, Trials | OR-combination of interim results. Limit to Cochrane Reviews, Protocols, Trials: final result |

| Results | Initial search<br>Search date 2018-10-29 | Update search<br>Publication Year 2018-2020<br>Search date 2020-09-11 |
|---|---|---|
| Cochrane Reviews | 19<br>Issue 10 of 12, October 2018 | 12<br>Issue 9 of 12, September 2020 |
| Cochrane Protocols | 1<br>Issue 10 of 12, October 2018 | 0<br>Issue 9 of 12, September 2020 |
| Trials | 416<br>Issue 9 of 12, September 2018 | 268<br>Issue 9 of 12, September 2020 |

ti,ab,kw = title,abstract keywords
near = terms in any order (default: within 6 words)
next = phrase searching: terms next to each other in the given order
* = truncation

Via Web of Science (Clarivate Analytics):
Databases:
**- Science Citation Index Expanded** (SCI-EXPANDED): 1945-present
**- Social Sciences Citation Index** (SSCI): 1956-present

| Set | Query | **Results Initial search** Search date: 2018-10-29 **Time span: all years** Data last updated: 2018-10-26 | **Results update search** Search date: 2020-09-11 **Timespan=2018-2020** Data last updated: 2020-09-10 | **Annotations** |
|---|---|---|---|---|
| # 1 | ts=(("data" near/3 share*) or ("data" near/3 sharing*)) | 12,398 | 5,227 | Concept data sharing |
| # 2 | ts=((patient* or participant*) near/3 individual*) | 73,929 | 17,281 | Concept Individual patient data sharing |
| # 3 | ts="data" | 5,101,598 | 1,087,419 | |
| # 4 | ts=(share* or sharing*) | 456,989 | 114,300 | |
| # 5 | #4 AND #3 AND #2 | 714 | 336 | |
| # 6 | ts=("IPD" near/6 (share* or sharing*)) | 23 | 24 | Concept IPD sharing |
| # 7 | #6 OR #5 OR #1 | 12,970 | 5,478 | OR-combination of concepts |
| # 8 | ts=(randomi?ed or "randomly" or randomi?ation) | 1,044,391 | 201,056 | Concept clinical trials |
| # 9 | ts=((random* or "clinical") near/3 trial*) | 773,198 | 154,885 | |
| # 10 | ts=("meta analy*" or metaanaly*) | 313,038 | 105,276 | Concept meta-analysis |
| # 11 | #10 or #9 OR #8 | 1,478,458 | 323,705 | Concept clinical trials OR meta-analysis |
| # 12 | #11 AND #7 | 1,022 | 453 | Concepts Data sharing AND (clinical trials OR meta-analysis) |
| # 13 | #11 AND #7 **Refined by: DOCUMENT TYPES:** ( ARTICLE OR REVIEW ) | 862 | 419 | Restriction to Article or Review: final result |

ts = topic: Title, Abstract, Author Keywords, Keywords Plus®
near/n = terms in any order within n words
* = truncation
? = wildcard for exact 1 character

# Supplementary material 3: Study characteristics

| Author | Year | Country | Type of research | Detail if type of research=other | Type of shared material |
|---|---|---|---|---|---|
| **Tudur-Smith C et al.** | 2014 | UK | Survey | | IPD |
| **Murugiah K et al.** | 2016 | US | Survey | | IPD |
| **Krleža-Jerić K et al.** | 2009 | Canada | Survey | | IPD |
| **Jones CW et al.** | 2016 | US | Survey | | IPD |
| **Mayo-Wilson E et al.** | 2015 | US | Other | Case study | IPD |
| **Reidpath DD et al.** | 2001 | Australia | Experim. | | IPD |
| **Chalmers I et al.** | 2013 | UK | Other | Case study | IPD |
| **Bergeris A et al.** | 2018 | US | Survey | | IPD |
| **Tudur Smith C et al.** | 2017 | UK | Other | Case study | Broader |
| **Vaduganathan M et al.** | 2018 | US | Metrics | | IPD |
| **Merson L et al.** | 2015 | Vietnam | Qualitative | | IPD |
| **Rowhani-Farid A et al.** | 2016 | Australia | Survey | | IPD |
| **Rathi V et al.** | 2012 | US | Survey | | IPD |
| **Ali J et al.** | 2015 | US | Survey | | IPD |
| **Hopkins C et al.** | 2016 | UK | Survey | | IPD |
| **Sydes M et al.** | 2015 | UK | Metrics | Case study | IPD |
| **Polanin J et al.** | 2019 | US | Experim. | | IPD |
| **Villain B et al.** | 2015 | France | Survey | | IPD |
| **Asare A et al.** | 2016 | US | Metrics | | IPD |
| **Strom B et al.** | 2016 | US | Other | Metrics + survey | IPD |
| **Mello M et al.** | 2005 | US | Survey | | IPD |
| **Rathi V et al.** | 2014 | US | Survey | | IPD |
| **Huser V et al.** | 2018 | US | Metrics | | IPD |
| **Chapman S et al.** | 2014 | UK | Survey | | IPD |
| **Griswold M et al.** | 2013 | US | Survey | | Broader |

| | | | | | |
|---|---|---|---|---|---|
| **Cheah PY et al.** | 2015 | Thailand | Qualitative | | IPD |
| **Hee SW et al.** | 2016 | UK | Other | Case study | IPD |
| **Geifman N et al.** | 2015 | US | Metrics | | IPD |
| **Strom B et al.** | 2014 | US | Metrics | | IPD |
| **Ross J et al.** | 2016 | US | Survey | | IPD |
| **Boutron I et al.** | 2016 | France | Survey | | Broader |
| **Vidal-Infer A et al.** | 2018 | Spain | Survey | | Broader |
| **Krumholz H et al.** | 2016 | US | Metrics | | IPD |
| **Tannenbaum S et al.** | 2018 | US | Survey | | IPD |
| **Ross J et al.** | 2017 | US | Metrics | | IPD |
| **Mello M et al.** | 2018 | US | Survey | | IPD |
| **Chickramane A et al.** | 2017 | India | Survey | | IPD |
| **Howe N et al.** | 2018 | UK | Qualitative | | IPD |
| **Naudet F et al.** | 2018 | US | Survey | | IPD |
| **Yuanyuan J et al.** | 2017 | China | Survey | | IPD |
| **Spence O et al.** | 2018 | US | Survey | | IPD |
| **Polanin J et al.** | 2018 | US | Survey | | IPD |
| **Zhu C et al.** | 2017 | US | Metrics | | IPD |
| **So D et al.** | 2017 | Canada | Survey | | IPD |
| **Savage C et al.** | 2009 | US | Survey | | IPD |
| **Kawahara T et al.** | 2018 | Japan | Metrics | | IPD |
| **Goldacre B et al.** | 2017 | UK | Survey | | IPD |
| **Pisani E et al.** | 2017 | UK | Other | Metrics + Qualitative research | IPD |
| **Bertagnolli M et al.** | 2017 | US | Metrics | | IPD |
| **Coady S et al.** | 2017 | US | Metrics | | IPD |
| **Hopkins A et al.** | 2018 | Australia | Survey | Survey | IPD |
| **Piwowar H et al.** | 2007 | US | Survey | | IPD |
| **Laine C et al.** | 2009 | US | Survey | | Broader |
| **Shmueli-Blumberg D et al.** | 2013 | US | Metrics | | IPD |
| **de Vito N et al.** | 2018 | UK | Survey | | Broader |

| | | | | |
|---|---|---|---|---|
| **Nevitt S et al.** | 2017 | UK | Survey | IPD |
| **Ahmed I et al.** | 2011 | UK | Survey | IPD |
| **Navar A et al.** | 2016 | USA | Metrics | IPD |
| **Ebrahim S et al.** | 2014 | USA | Survey | IPD |
| **Vassar M et al.** | 2020 | USA | Survey | IPD |
| **Cheah PY et al.** | 2018 | Thailand | Qualitative | IPD |
| **Staham EE et al.** | 2020 | USA | Survey | Broader |
| **Nutu D et al.** | 2019 | Romania | Survey | IPD |
| **Aleixandre-Benavent R et al.** | 2019 | Spain | Survey | IPD |
| **Ross JS et al.** | 2018 | USA | Metrics | Broader |
| **Gorman DM et al.** | 2019 | USA | Survey | IPD |
| **Bosserdt M et al.** | 2019 | Germany | Survey | IPD |
| **Whitlock EP et al.** | 2019 | USA | Survey | IPD |
| **Gabelica M et al.** | 2019 | Croatia | Survey | IPD |
| **Gorman DM et al.** | 2020 | USA | Survey | IPD |
| **Kaufmann I et al.** | 2019 | UK | Survey | IPD |
| **Veroniki AA et al.** | 2019 | Greece | Experim. | IPD |
| **Godolphin PJ et al.** | 2019 | UK | Experim. | Broader |
| **Rowhani-Farid A et al.** | 2020 | USA | Experim. | IPD |
| **Siebert M et al.** | 2020 | France | Survey | IPD |
| **Mayer C et al.** | 2019 | USA | Survey | IPD |
| **Gaba JF et al.** | 2020 | France | Survey | Broader |
| **Colombo C et al.** | 2019 | Italy | Survey | IPD |
| **Kochhar S et al.** | 2019 | India | Metrics | IPD |
| **Broes S et al.** | 2020 | Belgium | Qualitative | IPD |
| **Rollando P et al.** | 2020 | France | Survey | Broader |
| **Schmidt H et al.** | 2018 | Germany | Metrics | IPD |
| **Azar M et al.** | 2020 | Canada | Survey | IPD |
| **Almaqrami BS et al.** | 2020 | China | Survey | IPD |

| | | | | |
|---|---|---|---|---|
| **Papageorgiou SN et al.** | 2019 | Switzerland | Survey | iPD |
| **Miller J et al.** | 2019 | USA | Survey | Broader |
| **Lovato L et al.** | 2018 | USA | Metrics | IPD |
| **Kemper JM et al.** | 2020 | Australia | Survey | IPD |
| **Johnson AL et al.** | 2020 | USA | Survey | Broader |
| **Sherry C et al.** | 2019 | USA | Survey | Broader |
| **Pellen C et al.** | 2020 | France | Survey | Broader |
| **Danchev V et al.** | 2020 | USA | Survey | IPD |
| **Li R et al.** | 2020 | USA | Metrics | IPD |

Broader: the definition is not solely restricted to IPD and can cover other type of additional material (e.g. protocol, code, etc).

# Supplementary material 4: Published studies about YODA and CSDR

| Reference | Repository/ platform | No. of trials included in repository/ platform | No. of requests | No. of access to data | No. of publications | Date of assessment |
|---|---|---|---|---|---|---|
| Ross, 2017 | YODA | 189 | 73 | 50 | 2 | 6/2017 |
| Vadugan athan, 2018 | YODA | 537 | | 30 | 3 | 5/2017 |
| Strom, 2016 | CSDR | 237 | 177 | 144 | 1* | 11/2015 |

*based upon an investigator survey with 24 responses

# Supplementary material 5: Published outputs from YODA (up to 1st July 2019) and CSDR (up to 31 August 2019)

| Published outputs | Title | Platform used | Identification of the proposal | Type of study | Request date |
|---|---|---|---|---|---|
| **Allott EH et al. 2017** | Statin Use, Serum Lipids, and Prostate Inflammation in Men with a Negative Prostate Biopsy: Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Smoking Is Associated with Acute and Chronic Prostatic Inflammation: Results from the REDUCE Study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Branche BL et al. 2017** | Sleep Problems are Associated with Development and Progression of Lower Urinary Tract Symptoms: Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Vidal AC et al. 2016** | Racial differences in prostate inflammation: results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2016** | Does Prostate Size Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from the REDUCE Trial. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Simon RM et al. 2017** | Does Peak Urine Flow Rate Predict the Development of Incident Lower Urinary Tract Symptoms in Men with Mild to No Current Symptoms? Results from REDUCE. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Moreira DM et al. 2015** | Chronic baseline prostate inflammation is associated with lower tumor volume in men with prostate cancer on repeat biopsy: Results from the REDUCE study. | CSDR | 631 | Secondary analysis | 29/10/2013 |
| **Kent DM et al. 2016** | Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. | CSDR | 647 | Methodological | 29/10/2013 |
| **Baay M et al. 2017** | Background rates of disease in Latin American children from a rotavirus vaccine study. | CSDR | 651 | Secondary analysis | 11/03/2014 |
| **Le Noury J et al. 2015** | Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. | CSDR | 669 | Re-analysis | 27/01/2014 |
| **Nevitt SJ et al. 2017** | Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. | CSDR | 674 | Methodological | 15/05/2014 |
| **Nevitt SJ et al. 2017** | Antiepileptic drug monotherapy for epilepsy: a network meta-analysis of individual participant data. | CSDR | 674 | Meta-analysis | 15/05/2014 |

| | | | | | |
|---|---|---|---|---|---|
| **Forbess LJ et al. 2017** | Failure of a systemic lupus erythematosus response index developed from clinical trial data: lessons examined and learned. | CSDR | 911 | Secondary analysis | 25/07/2014 |
| **Dennis JM et al. 2018** | Evaluating associations between the benefits and risks of drug therapy in type 2 diabetes: a joint modeling approach. | CSDR | 930 | Secondary analysis | missing |
| **Dennis JM et al. 2018** | Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data. | CSDR | 930 | Secondary analysis | missing |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 945 | Secondary analysis | 23/04/2014 |
| **Mistry HB et al. 2017** | Model based analysis of the heterogeneity in the tumour size dynamics differentiates vemurafenib, dabrafenib and trametinib in metastatic melanoma. | CSDR | 946 | Secondary analysis | 28/05/2014 |
| **Muff S et al. 2018** | Bias away from the null due to miscounted outcomes? A case study on the TORCH trial. | CSDR | 977 | Re-analysis | 12/05/2014 |
| **Fragoso CAV et al. 2018** | Spirometric Criteria for Chronic Obstructive Pulmonary Disease in Clinical Trials of Pharmacotherapy. | CSDR | 993 | Secondary analysis | 28/02/2017 |
| **Devilliers H et al. 2016** | Minimal Clinically Important Differences for Generic Patient Reported Outcomes Tools in SLE | CSDR | 998 | Secondary analysis | missing |
| **Li-Kim-Moy J et al. 2018** | Impact of Fever and Antipyretic Use on Influenza Vaccine Immune Reponses in Children. | CSDR | 1000 | Secondary analysis | 08/09/2014 |
| **Blanco JR et al. 2017** | Impact of dolutegravir and efavirenz on immune recovery markers: results from a randomized clinical trial. | CSDR | 1028 | Secondary analysis | 23/09/2014 |
| **Borges NA et al. 2016** | Nonnucleoside Reverse-transcriptase Inhibitor- vs Ritonavir-boosted Protease Inhibitor-based Regimens for Initial Treatment of HIV Infection: A Systematic Review and Metaanalysis of Randomized Trials. | CSDR | 1058 | Meta-analysis | 18/08/2014 |
| **Dodd S et al. 2018** | Incidence and characteristics of the nocebo response from meta-analyses of the placebo arms of clinical trials of olanzapine for bipolar disorder. | CSDR | 1078 | Meta-analysis | 09/10/2014 |
| **Serrano-Villar S et al. 2017** | Effects of Maraviroc versus Efavirenz in Combination with Zidovudine-Lamivudine on the CD4/CD8 Ratio in Treatment-Naive HIV-Infected Individuals. | CSDR | 1079 | Secondary analysis | 12/10/2014 |
| **Emamikia S et al. 2017** | Relationship between glucocorticoid dose and adverse events in systemic lupus erythematosus: data from a randomized clinical trial. | CSDR | 1084 | Secondary analysis | 20/02/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Gruber JF et al. 2018** | Timing and predictors of severe rotavirus gastroenteritis among unvaccinated infants in low- and middle-income countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Gruber JF et al. 2018** | Timing of Rotavirus Vaccine Doses and Severe Rotavirus Gastroenteritis Among Vaccinated Infants in Low- and Middle-income Countries. | CSDR | 1088 | Secondary analysis | 04/09/2015 |
| **Schwartz LM et al. 2016** | Rotavirus vaccine effectiveness in low-income settings: An evaluation of the test-negative design. | CSDR | 1090 | Secondary analysis | 15/04/2015 |
| **Hilkens NA et al. 2016** | Blood pressure levels and the risk of intracerebral hemorrhage after ischemic stroke. | CSDR | 1100 | Secondary analysis | 13/01/2015 |
| **Hieronymus F et al. 2017** | Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. | CSDR | 1103 | Meta-analysis | missing |
| **Waljee AK et al. 2018** | Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. | CSDR | 1136 | Secondary analysis | 13/08/2015 |
| **Hadjichrysanthou C et al. 2016** | Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. | CSDR | 1137 | Secondary analysis | 16/04/2015 |
| **Voysey M et al. 2017** | The Influence of Maternally Derived Antibody and Infant Age at Vaccination on Infant Vaccine Responses : An Individual Participant Meta-analysis. | CSDR | 1141 | Meta-analysis | 22/07/2015 |
| **Radua J et al. 2017** | Meta-Analysis of the Risk of Subsequent Mood Episodes in Bipolar Disorder. | CSDR | 1148 | Meta-analysis | 30/01/2015 |
| **de Vries YA et al. 2018** | Initial severity and antidepressant efficacy for anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: An individual patient data meta-analysis. | CSDR | 1173 | Meta-analysis | 30/06/2015 |
| **Zafack JG et al. 2019** | Adverse events following immunisation with four-component meningococcal serogroup B vaccine (4CMenB): interaction with co-administration of routine infant vaccines and risk of recurrence in European randomised controlled trials. | CSDR | 1224 | Secondary analysis | missing |
| **Sturm A et al. 2017** | Evaluating the Hierarchical Structure of ADHD Symptoms and Invariance Across Age and Gender. | CSDR | 1292 | Methodological | 29/07/2015 |
| **Oon S et al. 2019** | Lupus Low Disease Activity State (LLDAS) discriminates responders in the BLISS-52 and BLISS-76 phase III trials of belimumab in systemic lupus erythematosus. | CSDR | 1320 | Secondary analysis | missing |
| **Craig K et al. 2017** | More of what works: Detection of informative sites during the conduct of clinical trials using machine learning | CSDR | 1323 | Methodological | 21/10/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Bauza C et al. 2018** | Determining the Joint Effect of Obesity and Diabetes on All-Cause Mortality and Cardiovascular-Related Mortality following an Ischemic Stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Bauza C et al. 2018** | Determining the joint effect of obesity and diabetes on functional disability at 3-months and on all-cause mortality at 1-year following an ischemic stroke. | CSDR | 1331 | Secondary analysis | 28/01/2016 |
| **Tajgardoon M et al. 2018** | A Novel Representation of Vaccine Efficacy Trial Datasets for Use in Computer Simulation of Vaccination Policy. | CSDR | 1374 | Secondary analysis | 25/05/2016 |
| **Berenguer J et al. 2019** | Mathematical modeling of HIV-1 transmission risk from condomless anal intercourse in HIV-infected MSM by the type of initial ART. | CSDR | 1403 | Secondary analysis | missing |
| **Hilkens NA et al. 2017** | Predicting Major Bleeding in Ischemic Stroke Patients With Atrial Fibrillation. | CSDR | 1455 | Secondary analysis | 03/06/2016 |
| **Kerr SJ et al. 2017** | The FDA snapshot algorithm may overestimate the efficacy of initial art | CSDR | 1456 | Methodological | missing |
| **Samara MT et al. 2017** | Initial symptom severity of bipolar I disorder and the efficacy of olanzapine: a meta-analysis of individual participant data from five placebo-controlled studies. | CSDR | 1457 | Meta-analysis | 08/06/2016 |
| **Hopkins AM et al. 2018** | Risk Factors for Severe Diarrhea with an Afatinib Treatment of Non-Small Cell Lung Cancer: A Pooled Analysis of Clinical Trials. | CSDR | 1475 | Meta-analysis | missing |
| **Peters EM et al. 2018** | Melancholic Symptoms in Bipolar II Depression and Responsiveness to Lamotrigine in an Exploratory Pilot Study. | CSDR | 1569 | Secondary analysis | 01/11/2016 |
| **de Vries YA et al. 2018** | Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: individual patient data meta-analysis. | CSDR | 1575 | Meta-analysis | 11/10/2016 |
| **Gemeinsamer Bundesausschuss, 2019** | Nutzenbewertungsverfahren zum Wirkstoff Sitagliptin | CSDR | 1593 | Secondary analysis | 04/11/2016 |
| **Hopkins AM et al. 2019** | Effect of early adverse events on response and survival outcomes of advanced melanoma patients treated with vemurafenib or vemurafenib plus cobimetinib: A pooled analysis of clinical trial data. | CSDR | 1599 | Meta-analysis | missing |
| **Carbon M et al. 2018** | Tardive dyskinesia risk with first- and second-generation antipsychotics in comparative randomized controlled trials: a meta-analysis. | CSDR | 1624 | Meta-analysis | missing |

| | | | | | |
|---|---|---|---|---|---|
| **Schwarzman LS et al. 2018** | The Association of Previous Prostate Biopsy Related Complications and the Type of Complication with Patient Compliance with Rebiopsy Scheme. | CSDR | 1626 | Secondary analysis | 16/12/2016 |
| **Voysey M et al. 2017** | Prevalence and decay of maternal pneumococcal and meningococcal antibodies: A meta-analysis of type-specific decay rates. | CSDR | 1629 | Meta-analysis | 26/09/2016 |
| **Shapiro W et al. 2018** | Salmeterol Combined with Fluticasone Reduces Exacerbations More Effectively in Chronic Bronchitis Associated with Chronic Obstructive Pulmonary Disease: A Post-hoc Analysis of the TORCH Trial | CSDR | 1640 | Secondary analysis | 28/02/2017 |
| **Parodis I et al. 2018** | Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Established organ damage reduces belimumab efficacy in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Parodis I et al. 2019** | Predictors of low disease activity and clinical remission following belimumab treatment in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hernández-Breijo B et al. 2019** | Antimalarial agents diminish while methotrexate, azathioprine and mycophenolic acid increase BAFF levels in systemic lupus erythematosus. | CSDR | 1695 | Secondary analysis | 21/09/2017 |
| **Hopkins AM et al. 2019** | Predictors of Long-Term Disease Control and Survival for HER2-Positive Advanced Breast Cancer Patients Treated With Pertuzumab, Trastuzumab, and Docetaxel. | CSDR | 1741 | Secondary analysis | missing |
| **Janciauskiene S et al. 2019** | Serum Levels of Alpha1-antitrypsin and Their Relationship With COPD in the General Spanish Population. | CSDR | 2084 | Secondary analysis | missing |
| **Storgaard H et al. 2016** | Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis. | YODA | 2014-0340 | Meta-analysis | 19/11/2014 |
| **Christian KE et al. 2019** | Gender Differences and Other Factors Associated with Weight Gain Following Initiation of Infliximab: A Post Hoc Analysis of Clinical Trials. | YODA | 2014-0334 | Meta-analysis | 26/11/2014 |
| **Wang R et al. 2018** | Comparative Efficacy of Tumor Necrosis Factor-α Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis. | YODA | 2014-0291 | Meta-analysis | 08/12/2014 |
| **Waljee AK et al. 2017** | External Validation of a Thiopurine Monitoring Algorithm on the SONIC Clinical Trial Dataset. | YODA | 2014-0401 | Secondary analysis | 20/01/2015 |
| **Mospan GA et al. 2017** | 5-Day versus 10-Day Course of Fluoroquinolones in Outpatient Males with a Urinary Tract Infection (UTI). | YODA | 2015-0514 | Secondary analysis | 26/05/2015 |

| | | | | | |
|---|---|---|---|---|---|
| **Singh S et al. 2018** | Impact of Obesity on Short- and Intermediate-Term Outcomes in Inflammatory Bowel Diseases: Pooled Analysis of Placebo Arms of Infliximab Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Singh S et al. 2018** | Obesity and Response to Infliximab in Patients with Inflammatory Bowel Diseases: Pooled Analysis of Individual Participant Data from Clinical Trials. | YODA | 2015-0612 | Meta-analysis | 20/10/2015 |
| **Spertus J et al. 2018** | Risk of weight gain for specific antipsychotic drugs: a meta-analysis. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Spertus J et al. 2019** | Bayesian Meta-analysis of Multiple Continuous Treatments with Individual Participant-Level Data: An Application to Antipsychotic Drugs. | YODA | 2015-0678 | Meta-analysis | 29/01/2016 |
| **Zou X et al. 2018** | The role of PANSS symptoms and adverse events in explaining the effects of paliperidone on social functioning: a causal mediation analysis approach. | YODA | 2016-0716 | Secondary analysis | 24/02/2016 |
| **World Health Organization 2017** | WHO report (appendix) | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Mbuagbaw L et al. 2019** | Outcomes of Bedaquiline Treatment in Patients with Multidrug-Resistant Tuberculosis. | YODA | 2016-0734 | Meta-analysis | 24/02/2016 |
| **Corbett M et al. 2017** | Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. | YODA | 2016-0897 | Meta-analysis | 19/05/2016 |
| **Gay HC et al. 2017** | Feasibility, Process, and Outcomes of Cardiovascular Clinical Trial Data Sharing: A Reproduction Analysis of the SMART-AF Trial. | YODA | 2016-0912 | Re-analysis | 07/06/2016 |
| **Schneider-Thoma J et al. 2018** | Second-generation antipsychotic drugs and short-term mortality: a systematic review and meta-analysis of placebo-controlled randomised controlled trials. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Schneider-Thoma J et al. 2019** | Second-generation antipsychotic drugs and short-term somatic serious adverse events: a systematic review and meta-analysis. | YODA | 2016-0880 | Meta-analysis | 17/06/2016 |
| **Teply BA et al. 2019** | Risk of development of visceral metastases subsequent to abiraterone vs placebo: An analysis of mode of radiographic progression in COU-AA-302. | YODA | 2016-1057 | Secondary analysis | 01/09/2016 |
| **Loubersac T et al. 2019** | Neutrophil-to-lymphocyte Ratio as a Predictive Marker of Response to Abiraterone Acetate: A Retrospective Analysis of the COU302 Study. | YODA | 2016-1103 | Secondary analysis | 23/11/2016 |
| **Martin LJ et al. 2018** | Identification of subgroups of metastatic castrate-resistant prostate cancer (mCRPC) patients treated with abiraterone | YODA | 2016-1122 | Secondary analysis | 23/11/2016 |

| | | | | | |
|---|---|---|---|---|---|
| | plus prednisone at low- vs. high-risk of radiographic progression: An analysis of COU-AA-302. | | | | |
| **Waljee AK et al. 2019** | Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. | YODA | 2016-1176 | Secondary analysis | 01/03/2017 |
| **Kubo K et al. 2018** | Placebo effects in adult and adolescent patients with schizophrenia: combined analysis of nine RCTs. | YODA | 2017-1676 | Meta-analysis | 24/05/2017 |
| **Kumagai F et al. 2018** | Early Placebo Improvement Is a Marker for Subsequent Placebo Response in Long-Acting Injectable Antipsychotic Trials for Schizophrenia: Combined Analysis of 4 RCTs. | YODA | 2017-1701 | Meta-analysis | 01/06/2017 |
| **Yiu ZZN et al. 2019** | A standardization approach to compare treatment safety and effectiveness outcomes between clinical trials and real-world populations in psoriasis. | YODA | 2017-1706 | Methodological | 08/08/2017 |
| **Narula N et al. 2018** | Patient-Reported Outcomes and Endoscopic Appearance of Ulcerative Colitis: A Systematic Review and Meta-analysis. | YODA | 2017-2031 | Meta-analysis | 30/08/2017 |
| **Singh S et al. 2018** | No Benefit of Concomitant 5-Aminosalicylates in Patients With Ulcerative Colitis Escalated to Biologic Therapy: Pooled Analysis of Individual Participant Data From Clinical Trials. | YODA | 2017-2306 | Meta-analysis | 25/09/2017 |
| **Singh S et al. 2019** | Efficacy and Speed of Induction of Remission of Infliximab vs Golimumab for Patients With Ulcerative Colitis, Based on Data From Clinical Trials. | YODA | 2018-3121 | Meta-analysis | 21/05/2018 |

# Status, use and impact of sharing Individual Participant data from clinical trials: a scoping review

C. Ohmann et al.

**Scoping Reviews (PRISMA.ScR) Checklist:**

| Section | Item | Covered | Page no. in manuscript |
|---|---|---|---|
| Title | Title | yes | 1 |
| Abstract | Structured summary | yes | 2 |
| Introduction | Rationale | yes | 3 |
| | Objectives | yes | 4 |
| Methods | Protocol and registration | yes | 5 |
| | Eligibility criteria | yes | 5 |
| | Information sources | yes | 6 |
| | Search | yes | 7 |
| | Selection of sources of evidence | yes | 7 |
| | Data charting process | yes | 7 |
| | Data items | yes | 7 |
| | Critical appraisal of individual sources of evidence | yes | 7 |
| | Synthesis of results | yes | 8 |
| Results | Selection of sources of evidence | yes | 8 |
| | Characteristics of sources of evidence | yes | 8 |
| | Critical appraisal within sources of evidence | yes | 9 |
| | Results of individual sources of evidence | yes | 9 |
| | Synthesis of results | yes | 9-20 |
| Discussion | Summary of evidence | yes | 20 |
| | Limitations | yes | 22 |
| | Conclusions | yes | 22 |
| Funding | Funding | yes | 3 |