



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

Christian Ohmann, David Moher, Maximilian Siebert, Edith Motschall,  
Florian Naudet

## ► To cite this version:

Christian Ohmann, David Moher, Maximilian Siebert, Edith Motschall, Florian Naudet. Status, use and impact of sharing individual participant data from clinical trials: a scoping review. *BMJ Open*, 2021, 11 (8), pp.e049228. 10.1136/bmjopen-2021-049228 . hal-03333562

**HAL Id: hal-03333562**

**<https://hal.science/hal-03333562>**

Submitted on 3 Sep 2021






**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# BMJ Open Status, use and impact of sharing individual participant data from clinical trials: a scoping review

Christian Ohmann <sup>1</sup>, David Moher <sup>2</sup>, Maximilian Siebert <sup>3</sup>,  
Edith Motschall <sup>4</sup>, Florian Naudet <sup>5</sup>

**To cite:** Ohmann C, Moher D, Siebert M, *et al.* Status, use and impact of sharing individual participant data from clinical trials: a scoping review. *BMJ Open* 2021;**11**:e049228. doi:10.1136/bmjopen-2021-049228

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online. (<http://dx.doi.org/10.1136/bmjopen-2021-049228>).

Received 20 January 2021

Accepted 20 July 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>European Clinical Research Infrastructure Network, Paris, France

<sup>2</sup>Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

<sup>3</sup>CHU Rennes, CIC 1414 (Centre d'Investigation Clinique de Rennes), University Rennes, Rennes, France

<sup>4</sup>Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Baden-Württemberg, Germany

<sup>5</sup>CHU Rennes, INSERM CIC 1414 (Centre d'Investigation Clinique de Rennes), University Rennes, Rennes, Bretagne, France

## Correspondence to

Professor Christian Ohmann;  
[christianohmann@outlook.de](mailto:christianohmann@outlook.de)

## 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 and 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 reuse 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.

## INTRODUCTION

### Rationale

Data sharing is increasingly recognised as a key requirement in clinical research.<sup>1</sup> In any

## 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.

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 (eg, 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 IPD 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.<sup>2</sup> 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,<sup>3</sup> to date data sharing is not the norm in clinical research, unlike many other scientific disciplines.<sup>4</sup> 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’.<sup>5</sup> 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 deidentified 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<sup>6</sup> 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.<sup>7</sup>

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.<sup>8</sup> Nevertheless, some innovative and progressive funders (eg, Wellcome Trust, Bill & Melinda Gates Foundation), and publishers/journals (eg, 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<sup>9</sup>—although often under controlled access conditions rather than simply being publicly available on-line<sup>10</sup>—and can structure that data to make it FAIR (findable, accessible, interoperable and reusable).<sup>11</sup> Meanwhile several data-sharing platforms and repositories are available and in use to provide practical support for the data-sharing process in clinical research (eg, 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 12 September 2018 (registration number: [osf.io/pb8cj](https://osf.io/pb8cj)). The protocol followed the methodology manual published by the Joanna Briggs Institute for scoping reviews.<sup>12</sup> Methods and results are reported using the Preferred Reporting Items for Systematic Reviews and Meta Analyses extension for Scoping Reviews.<sup>13</sup>

### 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 (eg, preprints, 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](#).

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’.<sup>14</sup> 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 (eg, on genomics) nor different fields outside medicine (eg, 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)

## Box 1 Definitions used for the five outcome domains

### 1. Intention to share data

There is an intention to share data, expressed by a stakeholder (eg, sponsor/principal investigator, 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 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)individual participant data 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 (eg, development of new hypotheses and methods) and/or medical health (eg, changes in treatment via guidelines).

and that promoted data-sharing practices (eg, at editorial level, at funder level, at research level). We considered many different types of studies (eg, experimental studies, surveys, metrics, quality assurance studies, qualitative research, reviews, reports), as the inclusion criteria were not method-specific but rather content-specific.

## Box 2 Definitions used for initiatives, repository and platform

### Initiatives

Major activities of an organisation (or a network of several organisations) to actively promote data-sharing in this area (eg, Pharmaceutical Research and Manufacturers of America/European Federation of Pharmaceutical Industries and Associations, Nordic Trial Alliance, Institute of Medicine, ICMJE, Research Data Alliance).

### 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 (eg, FreeBird, Biological Specimen and Data Repository Information Coordination Center or more general (eg, FigShare, Dryad).

### Platform

A computer environment where researchers can find datasets from clinical trials across different repositories, and where additional functionalities (eg, protected analysis environment) are provided (eg, ClinicalStudyDataRequest.com, Yale University Open Data Access, Project Data Sphere, Github).

## Information sources

The identification of studies was performed in two complementary stages:

- 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.
- Inspection of and if required contacts with known information sources (eg, 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 online supplemental material 1).

Between 25 January 2019 and 12 June 2019 (with an update on 2 November 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 online supplemental 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 online supplemental 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 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 October 2018 (update on 12 September 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 online supplemental material 2. The main search strategy developed by CO, DM and 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.<sup>15</sup> Discrepancies were resolved between the authors, and EM performed the search. All references were managed and deduplicated using a reference manager system (Endnote).



On 23 January 2019 (update on 2 November 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 January 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, programmes and material), (3) whether it reported data about one or more of the five outcomes domains defined in [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 reuse). We initially planned to describe the scale of reuse in qualitative terms and the observed results of the reuse (ie, '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 (eg, 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 prioritised 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 organised into (1) different sections corresponding to the key concepts detailed in the data-sharing pipeline (intention-to-share data, actual data sharing, results of reuse, 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 (eg, targeted group for intention to share data or type of use for reuse 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 summarised 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.

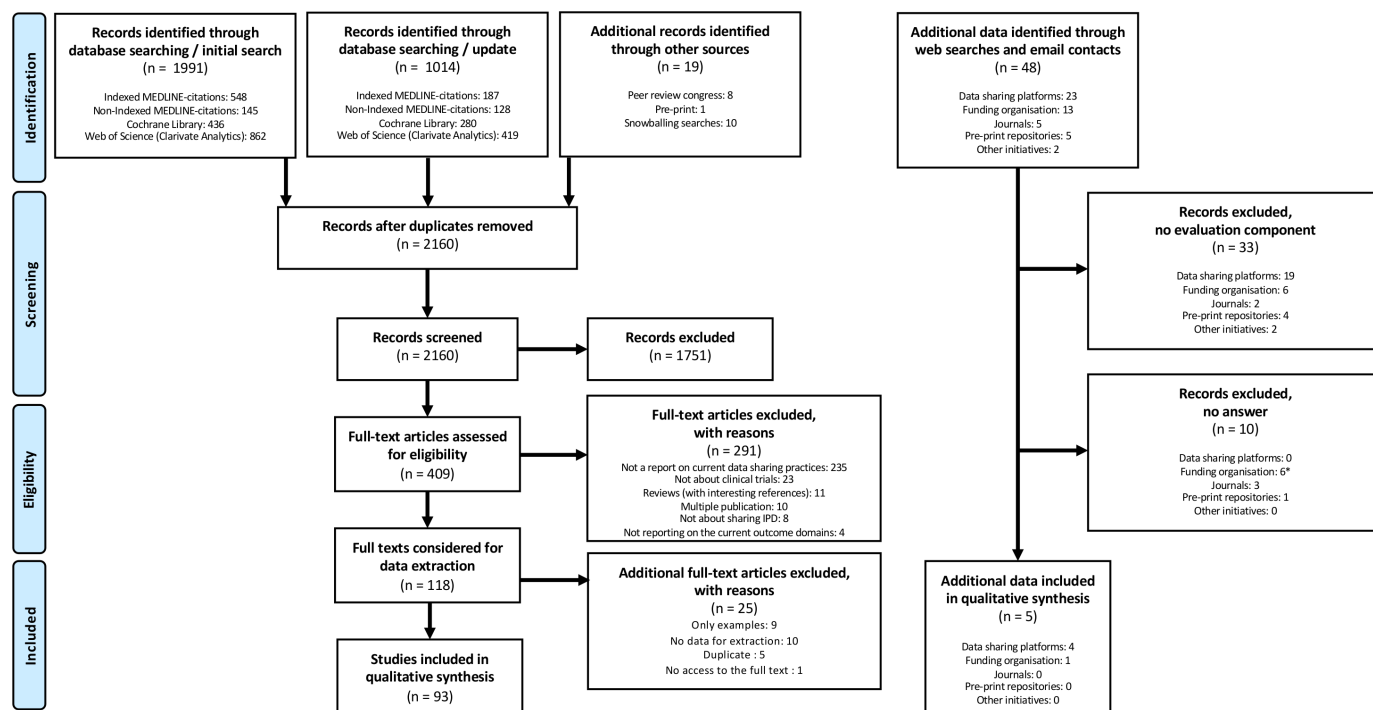
### Changes to the initial protocol

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, 3005 records (1991+1014 in the update) were retrieved by database search (2141 without duplicates). An additional eight 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 preprints. 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



**Figure 1** PRISMA flow diagram. \*For National Institute of Health USA, the answer we received was not informative. IPD, individual participant data; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

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. Four data-sharing platforms (CSDR, YODA, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Vivli) and one 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 and 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 (eg, 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 online supplemental 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 prespecified 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

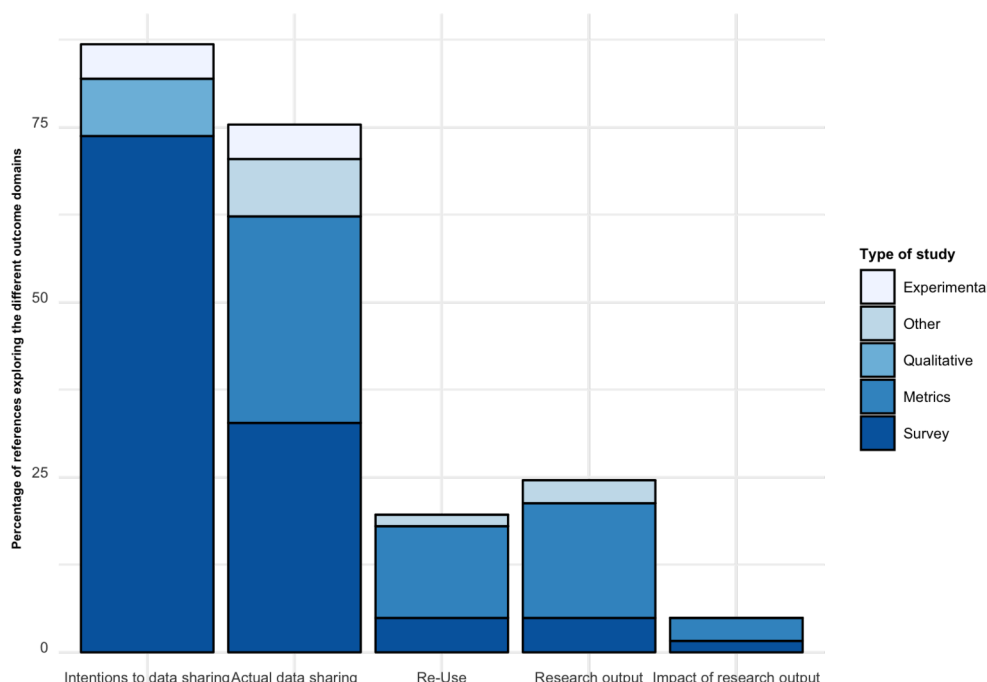
In general, there was a high risk of bias, especially due to study design (eg, surveys with low response rates and absence of experimental design). As stated in the methods, this was not assessed systematically. If available, we have tried to present this information in the narrative part of the review.

### 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%).<sup>16</sup> In the survey by Rathi *et al*, 74% and 72%, respectively, thought that sharing deidentified data through data



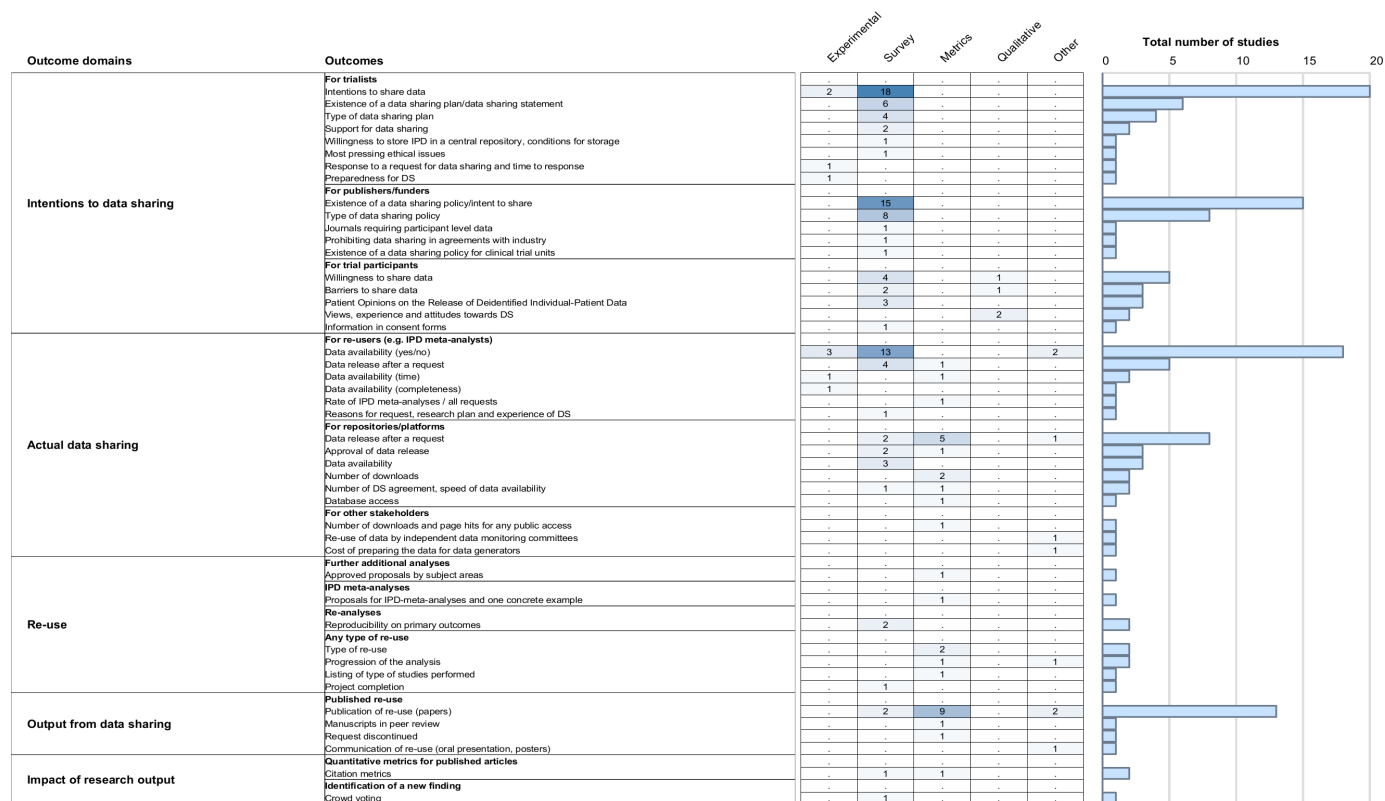
**Figure 2** Proportion of the 93 references exploring each outcome domain. Study designs considered. Experimental: prospective research that implies testing the impact a strategy (eg, randomised controlled trial). Survey: a general overview, exploration or description of individuals and/or research 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 (eg, case studies, environmental scans).

repositories should be required and that investigators should be required to share deidentified 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.<sup>17</sup> 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% vs 31%). The most academically productive trialists less frequently indicated they had withheld or would withhold data in order to protect research subjects (24 vs 40% for the least productive), as did those who had received industry funding compared with those who had not (24% vs 43%).<sup>18</sup> The survey by Tannenbaum, 2018 suggested that willingness to share data could depend on the intended reuse of the data (97% of respondents were willing to share data for a meta-analysis vs 73% for a reanalysis).<sup>19</sup> 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.<sup>20</sup>

#### *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 were available without condition for 4%, with conditions for 57% and unavailable for 38%.<sup>21</sup> Over the first 4 years, data were available without condition for 7%, with conditions for 47%, and unavailable for 46% of research articles.<sup>22</sup> Nine per cent 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.<sup>23</sup> Among 60 randomised cardiovascular interventional trials registered on ClinicalTrials.gov, up to 2015 with >5000 enrolment, 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.<sup>24</sup> Reasons for non-availability were: cosponsor did not agree to make IPD available (four trials) and trials were not conducted within a specific time (five trials); for the remaining six trials, no specific reason was provided. Of 619 RCTs published between 2014 and 2016 in seven 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.<sup>25</sup> 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



**Figure 3** Outcomes used to assess current data-sharing practices for individual patient data for clinical trials organised per outcome domain and number of studies exploring these outcomes. Study designs considered. Experimental: prospective research that implies testing the impact a strategy (eg, 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 (eg, case studies, environmental scans). IPD, individual participant data.

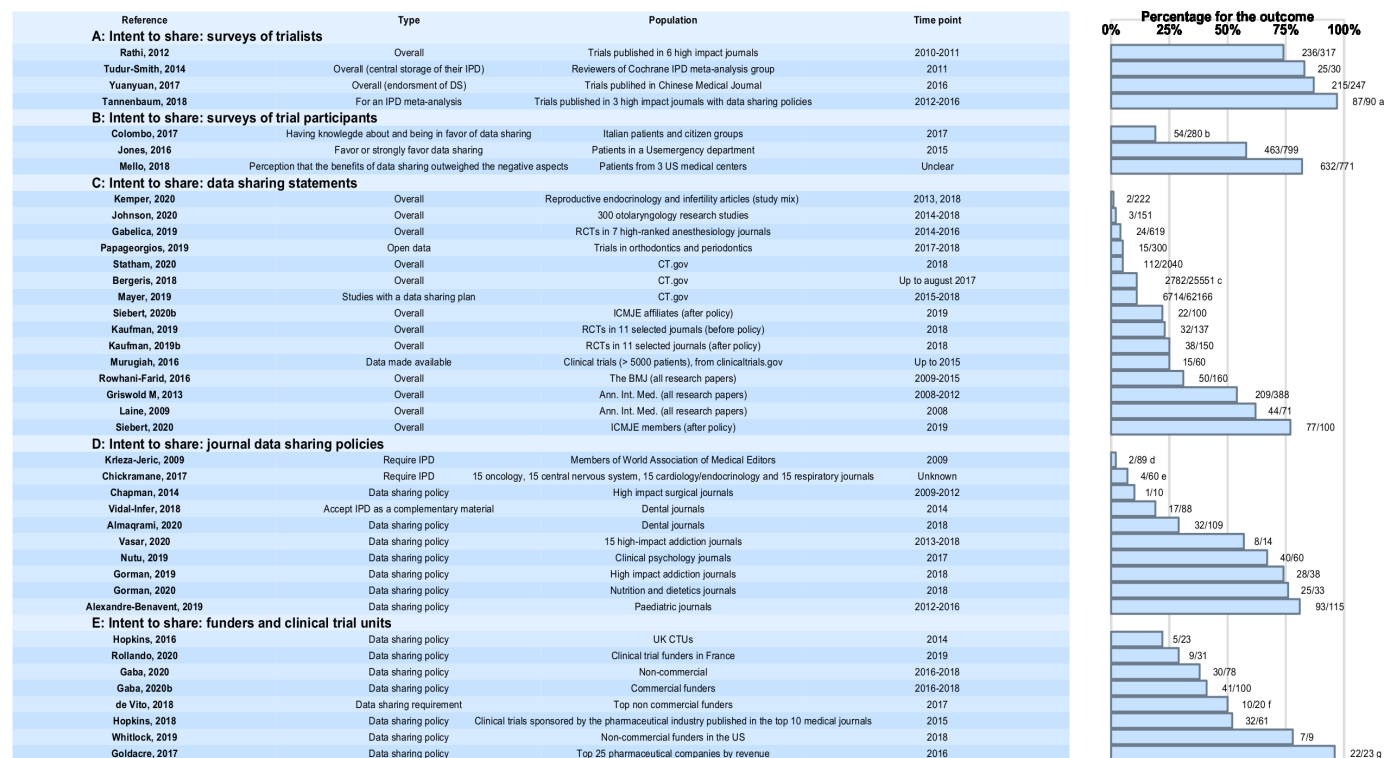
contain a data-sharing statement. The reporting of the statement was investigated in a 2-month period before and after this date.<sup>26</sup> The proportion of articles with a data-sharing statement was 23% (32/137) before and 25% (38/150) after 1 July 2018, while the number of journals publishing data-sharing statements increased from 4/11 to 7/11. Few data-sharing statements complied fully with the ICMJE journal criteria, and the majority did not refer to IPD. 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.<sup>27</sup> 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 and 2018.<sup>28</sup> Raw data were available on request or via online database for 1/98 articles in REI 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 five provided a data availability statement and 3 (2.0%) indicated that data were available.<sup>29</sup>

### 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'.<sup>30</sup> Differences were observed by key funder type, with 11% of National Institute of Health (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 2040 clinical trials first posted on ClinicaTrials.gov between 1 January 2018 and 6 June 2018 were investigated.<sup>31</sup> The vast majority of trials included in this study did not indicate an intention to share IPD (n=1928; 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





**Figure 4** Intent to share. Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference. <sup>a</sup>The proportion is 73% if the purpose is a reanalysis. <sup>b</sup>Fifty-four participants out of 60 had an opinion about data-sharing (the others had no knowledge or no opinion). <sup>c</sup>An additional 25% were undecided. <sup>d</sup>The proportion is 19% for requiring a data-sharing plan. <sup>e</sup>Thirty-five per cent have a data-sharing policy (encouraging data-sharing). <sup>f</sup>Only two with a mandatory policy. <sup>g</sup>The proportion is 71% for a sample of all companies (not only the top 25). In DeVito *et al*, we extracted the information on policies that made data-sharing mandatory (ie, a requirement to share the data). CTUs, Clinical Trial Units; ICMJE, International Committee of Medical Journal Editors; IPD, individual participant data; RCT, randomised controlled trial.

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.<sup>32</sup> Overall, 10.8% of trials with a first registration date after December 1 2015 answered 'yes' to plans to share deidentified 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 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.<sup>33</sup>

### Experimental studies

Experimental data suggest 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.<sup>34</sup> 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 reanalysis) and a specific request (a direct request for the data for reanalysis). Researchers receiving specific requests for data were less likely and slower to respond than researchers receiving general requests. Similarly, in 2019, a randomised 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.<sup>35</sup> The response rate was relatively low (21%) in this study since more than 1200 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 to 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').<sup>36</sup> Eighty-eight 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 nine 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, stigmatisation 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 (eg, trust in different types of research or organisations, relationships with the original research team).<sup>37</sup> Some qualitative reports provide data on heterogeneous samples including patients and various stakeholders from low-income 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.<sup>38</sup> A similar study on a range of relevant stakeholders in Thailand found that data sharing was seen as something positive (eg, 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.<sup>39</sup>

In a qualitative study with 16 in-depth interviews, patients with cancer currently participating in a clinical trial indicated a general willingness to allow reuse 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.<sup>40</sup> 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 USA 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 favoured or strongly favoured data sharing, while only 9% were against or strongly against it.<sup>41</sup> 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.<sup>42</sup> Predictors of this outcome were a low level of trust in others, concern about the risk of reidentification or about information theft, and having a college degree. Ninety-three per cent 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 enrol 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.<sup>43</sup> Of 280, 144 (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. Thirty-nine per cent 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 found that only 6 (4%) indicated a commitment to share deidentified IPD with third party researchers.<sup>44</sup> 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 (response rate 22%), 2% and 19% of journals, respectively, required provision of participant-level data and specification by authors of their data-sharing plan.<sup>45</sup> A similar survey of 10 high-impact surgical journals in 2009 and 2012 found only one journal that had a mandatory data-sharing policy.<sup>46</sup> 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).<sup>47</sup> A review of 88 websites of dental journals<sup>48</sup>

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.<sup>49</sup> Of 14, eight (57.1%) journals had data-sharing policies for published RCTs. Of the 394 RCTs included none shared their data publicly.

A total of 40/60 clinical psychology journals had a specific policy for data sharing (2017).<sup>50</sup> 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 six publication procedures, including data-sharing (2018). Of 38, 28 (74%) of the addiction journals had a data-sharing policy, none was mandatory.<sup>51</sup> 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).<sup>52</sup> Of 33, 25 (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.<sup>53</sup> 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.<sup>54</sup> Of 14, eight (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 IPD 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.<sup>55</sup> 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 USA. 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.<sup>56</sup> Trial transparency policies were investigated for 9/10 top non-commercial funders in the USA (May to November 2018).<sup>57</sup> Of nine, seven (78%) funders had a policy for individual patient data-sharing, for one it was mandatory. 6 offered data-sharing and five monitored compliance. Of 96 responders out of 190 non-commercial funders contacted in France, 31 were identified as funding clinical trials (2019).<sup>58</sup> Of 31, nine (29%) had implemented a data-sharing policy. Among these nine funders, only one had a mandatory sharing policy and eight 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. Fifty-two per cent 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.<sup>59</sup> A second survey studied data availability for 56 publications reporting on 61 industry-sponsored clinical trials of medications.<sup>60</sup> Of these 61 studies, 32 (52%) had a public data-sharing policy/process.

Seventy-eight 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).<sup>61</sup> Of 78, 30 (38%) non-commercial funders had a data-sharing policy with 18/30 (60%) making data-sharing mandatory and 12/30 (40%) encouraging data-sharing. Of 100, 41 (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 ClinicalTrials.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 were 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 Clinical Trial Units

Among 23 UK Clinical Research Collaboration registered Clinical Trial Units (CTUs) (response rate=51%), five (22%) had an established data-sharing policy and eight (35%) specifically required consent to use patient data beyond the scope of the original trial (see table).<sup>10</sup>



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 USA (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.<sup>62</sup> 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%).<sup>63</sup> 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.

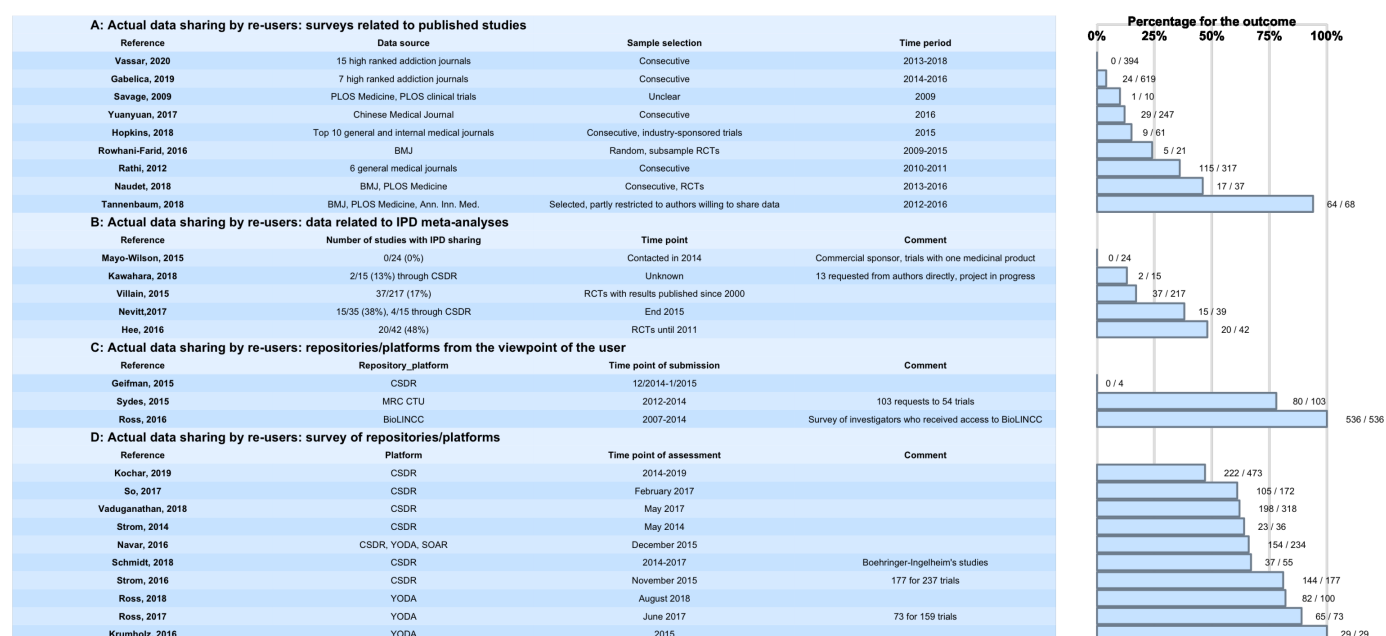
## Results for individual sources of evidence: actual data sharing Reusers

### 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.<sup>19</sup> 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.<sup>49</sup> Of 86 responders in a survey targeting the corresponding authors of 619 RCTs published between 2014 and 2016 in 7 high-ranking anaesthesiology journals, raw data were obtained only for 24 studies.<sup>25</sup> Sixty-two 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.<sup>64</sup> In articles in Chinese and international journals from 2016, sharing practices were indicated for 29/247 (11%) of the articles.<sup>20</sup> 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.<sup>60</sup> For BMJ research articles published between 2009 and 2015, data sets were made available in 7/157 (4%) of the articles.<sup>23</sup> For the subsample 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.<sup>17</sup> The data availability for RCTs published in BMJ and PLOS Medicine between 2013 and 2016 was 17/37 (46%).<sup>65</sup>



**Figure 5** Actual data sharing. Numbers correspond to the numbers of cases with the outcome/number of cases reported in each reference. CSDR, Clinical StudyDataRequest.com; CTU, Clinical Trial Unit; IPD, individual participant data; PLOS, Public Library of Science; SOAR, Supporting Open Access for Research Initiative; RCT, randomised controlled trial; YODA, Yale University Open Data Access.



### Experimental studies

In a parallel group RCT, an intervention group (offer of an Open Data Badge for data sharing) was compared with a control group (no badge for data sharing).<sup>66</sup> 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 were available on a third-party repository for 2/57 (3.5%) and on 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<sup>67</sup> and a study on antiepileptic drugs conducted by a Cochrane group with IPD for 15/39 (38%) studies included.<sup>68</sup> In another study on different databases, 35 IPD meta-analyses with more than 10 eligible RCTs were identified (1 May 2015 to 13 February 2017).<sup>69</sup> Of 774 eligible RCTs identified in these meta-analyses, 517 (66.8%) contributed data. The country where RCTs are conducted (the UK vs the USA), the impact factor of the journal (high vs 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.<sup>70</sup> Of 15 requests (13 direct to authors, 2 to a repository) in 2014/2016, IPD was received for 2/15 (13%) of the studies.<sup>71</sup> Of 217 RCTs published since 2000 in orthopaedic surgery, agreement to send IPD was obtained for 37/217 (17%).<sup>72</sup>

### 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 randomised prospective study where 29 corresponding authors of original research articles in a medical journal were contacted via two different modes (general vs 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.<sup>34</sup>

A randomised controlled trial investigated the effect of financial incentives on IPD sharing.<sup>73</sup> 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,<sup>74</sup> 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.<sup>68</sup> 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, Project Data Sphere (PDS)) were reported.<sup>75</sup> In these very early-phase projects, no data access was possible with CSDR, and faster data acquisition was achieved via the PDS. 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.<sup>76</sup> In a survey of investigators 536/536 (100%) received access to BioLINCC over a time period between 2007 and 2014.<sup>77</sup>

### 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).

Six 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.<sup>78</sup> Of these, 364 met initial administrative and data availability checks, and the independent review panel approved 291. Of 473, 222 (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.<sup>79</sup> 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.<sup>80</sup> 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.<sup>81</sup> From 2014 to 2017, Boehringer-Ingelheim listed 350 trials for potential data-sharing at CSDR.<sup>82</sup> Fifty-five 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.<sup>83</sup>

In the first year following the launch in October 2014, YODA received 29 requests all of which were approved (100%).<sup>84</sup> In 2017 the YODA project reported 73 proposals of which 65 were approved.<sup>85</sup> A more recent publication reported the metrics for data sharing of Johnson & Johnson clinical trials in the YODA project up to August 27, 2018.<sup>86</sup> One hundred data requests were received from 89 principal investigators (PI) for a median of 3 trials per request. Of 100, 90 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.<sup>87</sup>

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.

#### Metrics of trial coverage for data-sharing

Ethics approval in applications for open-access clinical trial data from CSDR was investigated in a survey.<sup>79</sup> 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.<sup>60</sup> 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).<sup>88</sup> 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.<sup>24</sup> 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 (eg, YODA, CSDR).<sup>89</sup> A total of 628 trials were analysed. Twenty-five per cent 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 (CTNs) 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 (NHLBI), access to 100 studies initiated between 1972 and 2010 was investigated.<sup>90</sup> 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.

**Table 1** Metrics of CSDR, YODA and Vivli websites

| Platform | Metrics date | Available studies | No of requests | No of requests with data shared | No of requests with data leading to publication | No of publications |
|----------|--------------|-------------------|----------------|---------------------------------|---|--------------------|
| CSDR     | 30/11/2020   | 3008              | 621            | 318                             | 59*   | 79                 |
| YODA     | 15/11/2019   | 334               | 196            | 173                             | 29  | 35                 |
| Vivli    | 2/11/2020    | 5203              | 215            | 123                             | 8   | 9                  |

NIDDK also provided metrics concerning the number of requests (530) but no other information.

\*Publication anticipated.

CSDR, ClinicalStudyDataRequest.com; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; YODA, Yale University Open Data Access.

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 CTN in the USA, with use increasing over the years.<sup>91</sup> 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.<sup>92</sup> 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 USA.

The 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.<sup>93</sup> 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 CTN.<sup>94</sup> 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 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.<sup>95</sup>

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

#### 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.<sup>97</sup> 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.<sup>98</sup> 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: reuse

##### Any type of reuse

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 optimising 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. Reanalyses are only exceptionally applied.

Early experiences with CSDR, involving GlaxoSmith-Kline trials found low rates of IPD meta-analyses and reanalyses, the vast majority being secondary analyses (studies on risk factors or biomarkers, methodological studies, predictive toxicology or risk models, studies of optimising treatments, subgroup analyses, etc).<sup>81</sup> Similar results were found in an update of the analysis.<sup>83</sup>

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).<sup>85</sup>

Among the 172 requests to the 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.<sup>90</sup> In only two requests, the available description suggested a reanalysis.

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).<sup>78</sup> Of 222, 90 (40.5%) of the research teams had completed their analyses by January 2018. Forty-one published at least one paper, and another 28 that were expected to publish shortly.

In the Systolic Blood Pressure Intervention Trial (SPRINT) challenge, individuals or groups were invited to analyse the dataset underlying the SPRINT RCT and to identify novel scientific or clinical findings.<sup>99</sup> 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, covering two large cancer screening trials (PLCO, NLST).<sup>96</sup> Of the 199 approved requests to PLCO between November 2012 and October 2016, 84 (42%) were devoted to cancer aetiology, 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 aetiology and 5 (2%) to risk prediction.

### IPD meta-analyses

In one study, IPD meta-analyses proved to amount to a small proportion of data reuse. Among the 174 research proposals approved up to 31 August 2017 by CSDR, 12 proposals were IPD meta-analyses, including network meta-analyses.<sup>71</sup> All were retrospective IPD meta-analyses (ie, none was a prospective IPD meta-analysis).

### Reanalyses

A 2014 survey of published reanalyses<sup>100</sup> found that a small number of reanalyses of RCTs have been published (only 37 reanalyses 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% CI: 59% to 94%) available studies were fully reproduced on all their primary outcomes.<sup>65</sup> 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 reuse of clinical trial data was considered in several studies. Detailed data are available for academic CTNs and disease-specific repositories in the USA, 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 5 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.<sup>77</sup> 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.<sup>63</sup> 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.

### Commercial sponsors

Various studies explored metrics of both YODA and CSDR (online supplemental material 4).

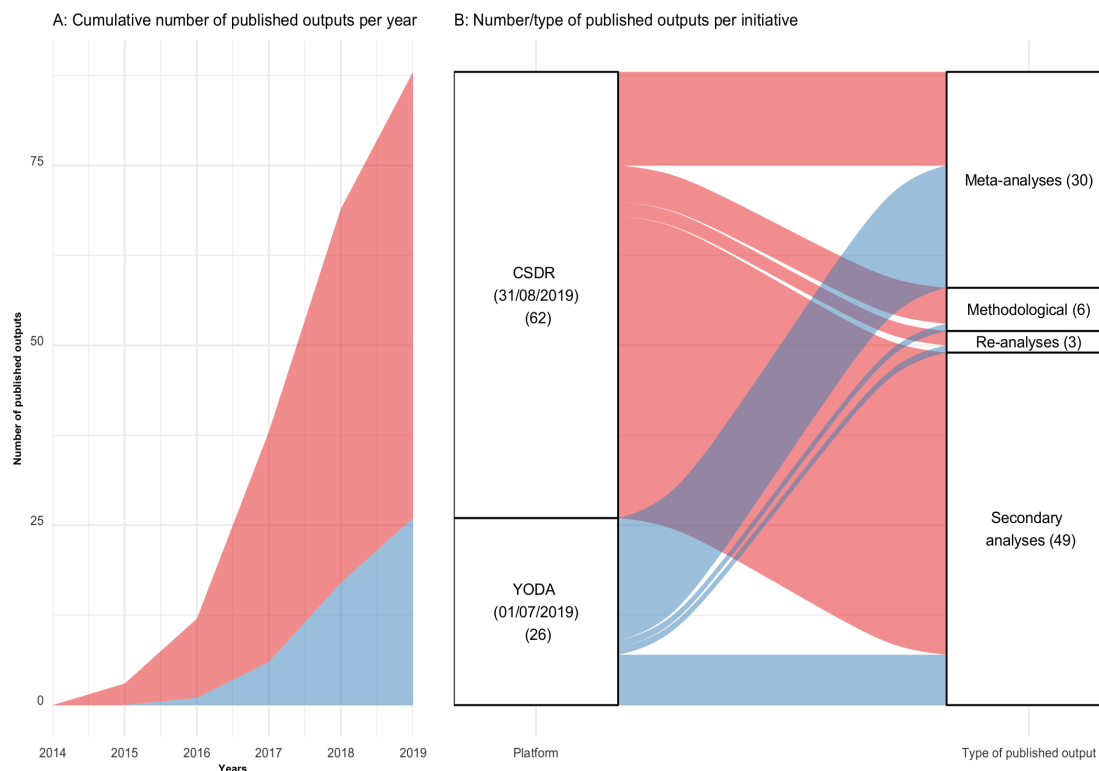
Up to 2021, Vivli's website indicates very little published output. We were not able to retrieve published output from NIDDK. Figure 6 presents publication metrics for CSDR (up to 31 August 2019) and YODA (up to 1 July 2019). Among 88 published papers (62 from CSDR and 26 from YODA), 49 were secondary 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

**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 <i>et al</i> , <sup>91</sup> | CTN Data Share      | 27 trials (1700 downloads)                   | 13   | 2012       |
| Zhu <i>et al</i> , <sup>96</sup>              | CDAS                | 2 trials (PLCO, NLST) (455 requests)         | 25% for PLCO projects, 19% for NLST projects   | 2016       |
| Coady <i>et al</i> , <sup>90</sup>            | BioLINCC            | 100 trials (88 requested at least once)      | 35% of clinical trials had at least 1 publication 5 years after availability in the repository | 5/2016     |
| Huser and Shmueli-Blumberg <sup>92</sup>      | NIDA Data Store     | 51 trials                                    | 14   | 3/2017     |
| Pisani and Botchway <sup>105</sup>            | WWARN               | 186 trials                                   | 18   | 2016       |

CDAS, Cancer Data Access system; CTN, clinical trial network; NIDA, National Institute of Drug Abuse; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal and Ovarian; WWARN, World Wide Antimalarial resistance Network.





**Figure 6** Temporal trends, number and type of published output from CSDR and YODA. Blue: YODA. Red: CSDR. CSDR, ClinicalStudyDataRequest.com; YODA, Yale University Open Data Access.

reanalyses (2 from CSDR and 1 from YODA). The details of these publications are presented in online supplemental material 5.<sup>80 83 85</sup>

### 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 were associated with an increased citation rate.<sup>101</sup> 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).<sup>90</sup> Half of the publications based on clinical trial data had cumulative citations that ranked in the top 34% normalised for subject

category and year of publication, compared with 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.<sup>99</sup> 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.

## 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 (eg, 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 (eg, *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 USA and NIH at the front. Only around 10%–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 (eg, 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 (eg, *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<sup>68 73</sup> and the time between request and receiving a data sharing agreement.<sup>75</sup>

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 optimising 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 repository (MDR), able to identify clinical studies and data objects related to it (eg, protocol, DMP, CRF).<sup>102</sup> This tool allows 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.<sup>2</sup> The situation could be considerably improved with more standardisation and harmonisation of data and procedures and a federating approach between repositories.

Reanalysis 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 reanalysis reached different conclusions with important implications for both clinical practice and research.<sup>3</sup> RIAT (Restoring invisible and abandoned trials support centre) 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.<sup>103</sup>

One of the problems that is tackled by RIAT is misreporting (inaccurately or incompletely reported trials). In our scoping review, we found that reanalyses 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 reanalysis 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 reuse, 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 CTNs and disease repositories have been successfully implemented (mainly in the USA) 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,<sup>77</sup> NIDA<sup>91</sup> and World Wide Antimalarial resistance Network.<sup>104 105</sup> This is reflected in the data-sharing rates for IPD meta-analyses, which are rather low if data requests target authors directly, compared with data-sharing requests within communities (eg, Cochrane groups) or related to specific repositories. Outside CTNs 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 (eg, 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 reanalysis is published. Existing approaches and tools could then be extended to automatically identify publications related to reuse of data and establish a link to the original work (eg, see crossmark—rossref,<sup>106</sup> MDR developed by ECRIN linking clinical studies with related data objects).<sup>102</sup> 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 (eg, generation of new hypotheses) and medical health (eg, changes in treatment via guidelines). Many interventions seek to maximise

the benefit of trial data sharing (eg, use of incentives for clinical trial data sharing, development of infrastructure for data sharing) 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 synthesising 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.<sup>107</sup> 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 reanalyse, synthesise, or build on previous data. First, a register (such as PROSPERO<sup>108</sup> 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.

**Twitter** David Moher @dmoher and Florian Naudet Florian Naudet @NaudetFlorian

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

**Contributors** 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.

**Funding** FN's work on data sharing is supported by a grant from the French National Research Agency-ANR (Reproducibility in Therapeutic Research/ReTheR: 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 no 654248). DM is supported by an Ottawa University Research Chair (grant number: N/A).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. All data relevant to the study are included in the article or uploaded as online supplemental information. There is a link to the code and the data on the OSF (<https://osf.io/h6cj4/>).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

## ORCID iDs

Christian Ohmann <http://orcid.org/0000-0002-5919-1003>

David Moher <http://orcid.org/0000-0003-2434-4206>

Maximilian Siebert <http://orcid.org/0000-0003-4385-5773>

Edith Motschall <http://orcid.org/0000-0001-7795-242X>

Florian Naudet <http://orcid.org/0000-0003-3760-3801>

## REFERENCES

- Nosek BA, Alter G, Banks GC, *et al*. Scientific standards. promoting an open research culture. *Science* 2015;348:1422–5.
- Banzi R, Canham S, Kuchinke W, *et al*. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials* 2019;20:169.
- 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.
- Data sharing and the future of science. *Nat Commun* 2018;9:2817.
- 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:e1001950.
- Rosenbaum L. Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp. *N Engl J Med* 2017;376:2201–3.
- 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:e1002315.
- Vasilevsky NA, Minnier J, Haendel MA, *et al*. Reproducible and reusable research: are Journal data sharing policies meeting the mark? *PeerJ* 2017;5:e3208.
- 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:e018647.
- 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. *J Clin Epidemiol* 2016;70:17–25.
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
- Peters MDJ, Godfrey CM, Khalil H, *et al*. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc* 2015;13:141–6.
- Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
- Clinicaltrials.gov. Learn about clinical studies. Available: <https://clinicaltrials.gov/ct2/about-studies/learn> [Accessed 21 Nov 2019].
- McGowan J, Sampson M, Salzwedel DM, *et al*. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016;75:40–6.
- 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:e97886.
- Rathi V, Dzara K, Gross CP, *et al*. Sharing of clinical trial data among Trialists: a cross sectional survey. *BMJ* 2012;345:e7570.
- 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.
- Tannenbaum S, Ross JS, Krumholz HM, *et al*. Early experiences with Journal data sharing policies: a survey of published clinical trial Investigators. *Ann Intern Med* 2018;169:586–8.
- International Congress on Peer Review and Scientific Publication. 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*, Chicago, 2017.
- International Congress on Peer Review and Biomedical Publication. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. ; 2009; Vancouver. *The 6th International Congress on Peer Review and Biomedical Publication*, Vancouver, 2009.
- Griswold M, Laine C, Mulrow C. 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*, 2013.
- Rowhani-Farid A, Barnett AG. Has open data arrived at the British medical Journal (BMJ)? an observational study. *BMJ Open* 2016;6:e011784.
- Murugiah K, Ritchie JD, Desai NR, *et al*. Availability of clinical trial data from industry-sponsored cardiovascular trials. *J Am Heart Assoc* 2016;5:e003307.
- Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share RAW data. *J Clin Epidemiol* 2019;109:111–6.
- 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 ISMP. *Current Medical Research and Opinion* 2019;35:31–41.
- Papageorgiou SN, Antonoglou GN, Martin C, *et al*. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. *J Orthod* 2019;46:101–9.
- Kemper JM, Rolnik DL, Mol BWJ, *et al*. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. *Fertil Steril* 2020;114:1322–9.
- Johnson AL, Torgerson T, Skinner M, *et al*. An assessment of transparency and reproducibility-related research practices in otolaryngology. *Laryngoscope* 2020;130:1894–901.
- Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at ClinicalTrials.gov. *JAMA* 2018;319:406–8.



- 31 Statham EE, White SA, Sonwane B, *et al.* Primed to comply: individual participant data sharing statements on ClinicalTrials.gov. *PLoS One* 2020;15:e0226143.
- 32 Mayer CS, Williams N, Gold S, *et al.* Sharing of individual participant data from clinical trials: general comparison and HIV use case. *AMIA Annu Symp Proc* 2019;2019:647–54.
- 33 Ali J, Califf R, Sugarman J. Anticipated ethics and regulatory challenges in PCORnet: the National patient-centered clinical research network. *Account Res* 2016;23:79–96.
- 34 Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15:125–34.
- 35 Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *J Clin Epidemiol* 2019;106:60–9.
- 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:e025273.
- 37 Howe N, Giles E, Newbury-Birch D, *et al.* Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *J Health Serv Res Policy* 2018;23:123–33.
- 38 Merson L, Phong TV, Nhan LNT, *et al.* Trust, respect, and reciprocity: informing culturally appropriate Data-Sharing practice in Vietnam. *J Empir Res Hum Res Ethics* 2015;10:251–63.
- 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. *J Empir Res Hum Res Ethics* 2015;10:278–89.
- 40 Broes S, Verbaander C, Casteels M, *et al.* Sharing of clinical trial data and samples: the cancer patient perspective. *Front Med* 2020;7:33.
- 41 Jones CW, Roberts BW, Platts-Mills TF. Patient perspectives on sharing Deidentified trial data. *Ann Intern Med* 2016;165:748–9.
- 42 Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *N Engl J Med* 2018;378:2202–11.
- 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:e024863.
- 44 Spence O'Mareen, Onwuchekwa Uba R, Shin S, *et al.* Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials* 2018;19:269.
- 45 Karmela Krleza-Jerić, Ida Sim, Ana Marušić, Ludovic Reveiz and Carlos Granados. Medical Journal editor perspectives on sharing results data from patient-oriented research: a WAME survey. *6th Peer Review Congress*, Vancouver, 2009.
- 46 ASIT. *Promoting transparency in clinical research: systematic review of disclosure and data-sharing policies in surgical journals*, 2014.
- 47 Cochrane Library. 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. *Clin Oral Investig* 2018;22:275–80.
- 49 Vassar M, Jellison S, Wendelbo H, *et al.* Data sharing practices in randomized trials of addiction interventions. *Addict Behav* 2020;102:106193.
- 50 Nutu D, Gentili C, Naudet F, *et al.* Open science practices in clinical psychology journals: an audit study. *J Abnorm Psychol* 2019;128:510–6.
- 51 Gorman DM. Use of publication procedures to improve research integrity by addiction journals. *Addiction* 2019;114:1478–86.
- 52 Gorman DM, Ferdinand AO. High impact nutrition and dietetics journals' use of publication procedures to increase research transparency. *Res Integr Peer Rev* 2020;5:12.
- 53 Almqami BS, Hua F, Liu Y, *et al.* Research waste-related editorial policies of leading dental journals: situation 2018. *Oral Dis* 2020;26:696–706.
- 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:e038887.
- 55 Alexandre-Benavent R, Sapena AF, Ferrer SC, *et al.* Policies regarding public availability of published research data in pediatrics journals. *Scientometrics* 2019;118:439–51.
- 56 DeVito NJ, French L, Goldacre B. Noncommercial Funders' policies on trial registration, access to summary results, and individual patient data availability. *JAMA* 2018;319:1721–3.
- 57 Whitlock EP, Dunham KM, DiGiovio K, *et al.* Noncommercial us Funders' policies on trial registration, access to summary results, and individual patient data availability. *JAMA Netw Open* 2019;2:e187498.
- 58 Rollando P, Parc C, Naudet F, *et al.* [Data sharing policies of clinical trials funders in France]. *Therapie* 2020;75:527–36.
- 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.
- 60 Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. *BMC Med* 2018;16:165.
- 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:e0237464.
- 62 Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005;352:2202–10.
- 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. *Eur Radiol* 2019;29:4794–802.
- 64 Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* 2009;4:e7078.
- 65 Naudet F, Sakarovich 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.
- 66 Rowhani-Farid A, Aldcroft A, Barnett AG. Did awarding badges increase data sharing in *BMJ Open*? A randomized controlled trial. *R Soc Open Sci* 2020;7:191818.
- 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. *Eur J Pain* 2017;21:815–26.
- 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.
- 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. *J Clin Epidemiol* 2020;124:16–23.
- 70 Mayo-Wilson E, Doshi P, Dickersin K. Are manufacturers sharing data as promised? *BMJ* 2015;351:h4169.
- 71 Kawahara T, Fukuda M, Oba K, *et al.* Meta-Analysis of randomized clinical trials in the era of individual patient data sharing. *Int J Clin Oncol* 2018;23:403–9.
- 72 Villain B, Dechartres A, Boyer P, *et al.* Feasibility of individual patient data meta-analyses in orthopaedic surgery. *BMC Med* 2015;13:131.
- 73 Veroniki AA, Ashoor HM, Le SPC, *et al.* Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *J Clin Epidemiol* 2019;113:176–88.
- 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.
- 75 Geifman N, Bollyky J, Bhattacharya S, *et al.* Opening clinical trial data: are the voluntary data-sharing portals enough? *BMC Med* 2015;13:280.
- 76 Sydes MR, Johnson AL, Meredith SK, *et al.* Sharing data from clinical trials: the rationale for a controlled access approach. *Trials* 2015;16:104.
- 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:e012769.
- 78 Kochhar S, Knoppers B, Gamble C, *et al.* Clinical trial data sharing: here's the challenge. *BMJ Open* 2019;9:e032334.
- 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:e0184491.
- 80 Vaduganathan M, Nagarur A, Qamar A, *et al.* Availability and use of shared data from cardiometabolic clinical trials. *Circulation* 2018;137:938–47.
- 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;371:2052–4.
- 82 Schmidt H, Caldwell B, Mollet A, *et al.* An industry experience with data sharing. *N Engl J Med* 2018;379:2081–2.
- 83 Strom BL, Buyse ME, Hughes J, *et al.* Data Sharing - Is the Juice Worth the Squeeze? *N Engl J Med* 2016;375:1608–9.

- 84 Krumholz HM, Waldstreicher J. The Yale open data access (yodA) project — a mechanism for data sharing. *N Engl J Med Overseas Ed* 2016;375:403–5.
- 85 Ross JS, Ritchie JD, Bamford S. Sharing data through the Yale university open data access (yodA) project: early experience. *The 8th International Congress on Peer Review and Biomedical Publication*, 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. *Sci Data* 2018;5:180268.
- 87 Navar AM, Pencina MJ, Rymer JA, *et al*. Use of open access platforms for clinical trial data. *JAMA* 2016;315:1283–4.
- 88 Boutron I, Dechartres A, Baron G, *et al*. Sharing of data from Industry-Funded registered clinical trials. *JAMA* 2016;315:2729–30.
- 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.
- 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;376:1849–58.
- 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. *Clin Trials* 2013;10:977–86.
- 92 Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. *Clin Trials* 2018;15:413–23.
- 93 Bertagnolli MM, Sartor O, Chabner BA, *et al*. Advantages of a truly open-access Data-Sharing model. *N Engl J Med* 2017;376:1178–81.
- 94 Asare AL, Carey VJ, Rotrosen D, *et al*. Clinical trial data access: opening doors with TrialShare. *J Allergy Clin Immunol* 2016;138:724–6.
- 95 MRC. MRC clinical trials transparency review, 2017. Available: <https://mrc.ukri.org/documents/pdf/clinical-trials-review-2017/> [Accessed 22 Nov 2019].
- 96 Zhu CS, Pinsky PF, Moler JE, *et al*. Data sharing in clinical trials: an experience with two large cancer screening trials. *PLoS Med* 2017;14:e1002304.
- 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.
- 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;18:319.
- 99 Lovato L. Exploring the potential of data sharing: the sprint challenge in Abstracts from the SCT 39th annual meeting (2018). *Clinical trials* 2018;15:S35–192.
- 100 Ebrahim S, Sohani ZN, Montoya L, *et al*. Reanalyses of randomized clinical trial data. *JAMA* 2014;312:1024–32.
- 101 Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLoS One* 2007;2:e308.
- 102 ECRIN. ECRIN metadata Repository, 2021. Available: <https://ecrin.org/tools/clinical-research-metadata-repository> [Accessed 1 Jul 2021].
- 103 Doshi P, Shamseer L, Jones M, *et al*. Restoring biomedical literature with RIAT. *BMJ* 2018;361:k1742.
- 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. *Am J Trop Med Hyg* 2020;103:359–68.
- 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.
- 106 Crossmark. Available: <https://www.crossref.org/services/crossmark/> [Accessed 1 Nov 2021].
- 107 International Consortium of Investigators for Fairness in Trial Data Sharing, Devereaux PJ, Guyatt G, *et al*. Toward fairness in data sharing. *N Engl J Med* 2016;375:405–7.
- 108 Booth A, Clarke M, Dooley G, *et al*. The nuts and bolts of Prospero: an international prospective register of systematic reviews. *Syst Rev* 2012;1:2.