

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Open Sharing of Clinical Trial Data

Vera J. Lipton

IntechOpen

This chapter summarises the experience with the digital sharing of clinical trial data, focusing on the methods and stages of data sharing and the issues that act as barriers to data sharing.

The discussion is structured in four key sections:

1. The value of open clinical trial data
2. Stakeholders in the sharing of clinical trial data
3. The stages of data sharing
4. The challenges of open data sharing

Introduction

The previous chapter examined the practices and methods for sharing particle physics data and identified the key challenges and lessons learnt in the process. This chapter turns to consider the practice of the digital sharing of clinical trial data. Clinical testing of new pharmaceuticals is governed by rigorous ethics and research protocols.

In this chapter, I examine how these protocols are being applied to shared data resulting from clinical trials. The examination starts by identifying the drivers for the open sharing of clinical trial data. This is followed by discussion of the role of various stakeholders in driving the data release, focusing especially on pharmaceutical regulatory agencies. I then describe the stages of data release as they have emerged in recent years and compare these stages to those at CERN. The final section discusses the challenges arising in open sharing of clinical trial data. I pay particular attention to privacy concerns—to managing the risks of data misinterpretation and to the motivations for researchers to curate and to share data and collaborate with colleagues.

6.1 The value of open clinical trial data

Clinical trials are crucial in determining the safety and efficacy of pharmaceuticals and for ensuring appropriate and effective treatment. Clinical trials represent the largest portion of the estimated US\$1.3 billion total cost of developing a new drug and bringing it to the market [311]. In the digital era, doctors and patients increasingly require access to clinical trial data as they use the Internet and online resources to learn about diseases, treatment and side effects. Patients also share knowledge and experiences and participate in online communities, many of which are disease-specific. Such active involvement by patients can be helpful in

determining the course of their treatment and care. According to the President of the Institute for Clinical Systems Improvement:

An important component of patient contribution is to balance what evidence-based medicine and conventional medical wisdom recommends with what is possible, desirable and most acceptable for the individual patient.¹

In this emergent field of consumer-driven research, patients are no longer just passive recipients of care. They are becoming more aware and more active—becoming members of the care and research team, engaged in decision-making regarding their treatment plans.² At this time, such participation remains limited to highly educated patients and social groups [313–317], even though with increasing education levels around the world, the prospects for patient-driven medical research are increasing, too. Some online communities of patients have already made a difference to the quality of life of those suffering from rare cancers and other diseases [318, 319]. In some cases, patient participation in medical research has resulted in the ability to recruit participants for clinical trials in rare diseases and to get them actively involved in community engagement activities, as reported in the inaugural issue of the *Journal of Participatory Medicine*.³

As medical research is becoming more patient-driven, calls for broader access to clinical trial data intensify. The established system of data sharing is being tested. Access to clinical trial data has traditionally been made available only to researchers following the requests made to investigators, research sponsors or journal editors [320]. These traditional methods used for clinical trial data dissemination are far more advanced than in other scientific fields. For example, the descriptors for clinical datasets are more advanced than in other disciplines [321]. However, the sharing of clinical trial data still occurs in closed professional circles through direct sharing rather than online, via public repositories.

There are several reasons for such restricted data sharing. Firstly, many clinical trials are not reported [322]. Secondly, only positive results usually get published in journals [323, 324]. Thirdly, of those published results, only a small portion of clinical trial data gets deposited along with articles in peer-reviewed journals [325, 326]. Finally, the data that is supposed to get published under open data mandates may not be readily available to other users.

For example, a 2009 study of data-sharing requests made to investigators listed in *PLOS Medicine* and *PLOS Clinical Trials*—two authoritative journals committed to data sharing—found that there was a very limited compliance with the PLOS data-sharing policy. The policy explicitly stated that:

... data should be provided as supporting information with the published paper. If this is not practical, data should be made freely available upon reasonable requests.⁴

Upon lodging 10 such requests, the authors of the study were able to obtain only one dataset. Two email addresses of the investigators listed as contacts for data sharing were no longer valid. Four investigators were unwilling to share their data either because they were too busy or that it would take too much work to organise

¹ Comments made by Dr. Kent Bottles, President of The Institute for Clinical Systems Improvement. See Frydman [312].

² *Ibid.*

³ Frydman [312] at point 2.

⁴ *PLoS Medicine Editorial and Publishing Policies* as valid at 4 March, (2009). Cited in Savage and Vickers [327].

and to annotate the datasets or that data sharing was not yet possible because further analyses were needed with the data. Two investigators did not respond to the data requests and numerous follow-up emails. The authors of the study concluded that policies set by journals that require the sharing of data do not facilitate availability of datasets to independent investigators.⁵

A similar study published in 2006 summarised the results of data requests made directly to authors of articles published in psychology journals. Their success in obtaining data was slightly higher, even though only a handful of the journals had in place a data-sharing policy. Out of 141 requests, the authors received data in 26% of cases over the 6 months of their efforts to contact the authors [328].

The most recent and comprehensive study was published in 2013. That study, employing several research methods, examined the potential for sharing genomic data published under open data mandates. In one approach they directly contacted the authors of eligible papers from two journals committed to data publication. Specifically, they contacted 19 authors and achieved a 59% return rate within 7.7 days [329]. This result is markedly better than those reported in earlier studies. The authors surmised that the attitudes to data sharing have improved in recent years. However, they also noted that recent studies based on human data still had low success rates, perhaps due to privacy issues that present a barrier to data sharing in these scientific disciplines.⁶

The same study also examined the elements of open data policies that positively correlate with data publication. Based on extensive statistical analyses, the authors concluded that mandatory data-sharing policies can effectively drive data release, provided two additional conditions are met.

Firstly, the mandates should require the archiving of data in repositories and, secondly, that a data accessibility statement must be included in the manuscript. Mandatory data archiving does not itself stimulate availability of open data if the data accessibility statement in the manuscript is missing—all three requirements must be met. In other words, the most stringent policies comprising the three elements lead to significantly more data release and by some distance.⁷

A similar conclusion was reached by a 2011 study in the field of gene expression data (microarray datasets). Piwowar and Chapman examined whether correlation exists between the strength of open data mandates and the probability of data release. The authors found that only 17 out of 70 relevant scientific journals had in place 'strong' (meaning mandatory) data archiving policies. And those with strong open data mandates were more likely to achieve publication of research data, with a success rate of more than 50%. Journals with 'recommended archiving' achieved a deposit rate of just over 30%, while a journal with no policy had only about 20% availability [330].

Another interesting finding was that journals with a higher impact factor were more likely to have the data published online than journals with substantially lower impact factor. Accordingly, a positive correlation exists between high-quality research and the availability of data.⁸ A similar finding resulted from the 2013 study by Vines et al. [329].⁹

⁵ *Ibid.*

⁶ *Ibid.*

⁷ *Ibid.*, 36. In 'mandate archiving' into those journals for which a data accessibility statement was required in the manuscript, the chances of finding the data online were 974 times higher for those with no such policy.

⁸ *Ibid.* Specifically, Piwowar found that a 244 journal with 244 an impact factor (IF) of 15 was 4.5 times more likely to have the microarray data online than a 245 journal with an IF of 5.

⁹ See Vines et al. [329] at point 15.

From the studies discussed above, it is apparent that online sharing of clinical data is a desirable practice to enshrine in open data policies, even though compliance with those policies remains limited. The funders' and publishers' policies that merely encourage the deposit of research data generally achieve lower deposit rates than those with more prescriptive policies. In all cases, however, the deposit rates remain low and rarely exceed 50%. The lack of implementation suggests there are cultural and other factors that act as barriers to data release and open sharing.

Factors contributing to limited data sharing are discussed further in the following sections. The first section outlines the various stakeholders involved in the sharing of clinical trial data. This is followed by an explanation of the various forms and stages of data sharing and factors that have been identified as limiting or inhibiting researchers and their organisations from sharing research data.

6.2 Stakeholders in the sharing of clinical trial data

Several stakeholders are involved in the process of clinical trial data collection, processing, analysing, and subsequent sharing with other parties. These include the patients or other people participating in clinical trials, funders and other sponsors of trials, pharmaceutical regulatory agencies, medical research institutes and universities, external research investigators, journal publishers, and learned societies. The stakeholders play different roles in the process leading to data sharing.

6.2.1 Patients or other research subjects participating in trials

The role of human subjects participating in clinical trials has traditionally been limited to providing the necessary data. Patients involved in clinical trials were seen merely as data providers. In light of the calls for patient-centred health care, patient advocacy groups focus on the greater engagement of patients—not just in clinical trials, but also in the decisions made around the process of the design and conduct of clinical studies, including the process of data sharing [331–333]. Patient engagement increases study enrolment rates, improves credibility of the results, and assists researchers to secure funding and to design study protocols [334].

Patient involvement comes at increased costs, especially logistic costs. Yet, since much of the cost is borne by the patients themselves, the argument that the role of patients should be limited to just data provision is no longer plausible. At the same time, there are cultural barriers to greater patient involvement in the healthcare system. This is because both the established research method and societal expectation are to perform research *on* patients, and not *with* patients [335]. Accordingly, patients still continue to be regarded as a source of clinical trial data, rather than as active participants in the clinical research process [336].

With the emergence of online platforms promoting the engagement of patients in clinical trials, as summarised in **Table 3**,¹⁰ this practice may be changing. The impact of these platforms on the practice of clinical trials is yet to be seen.

The process of recruiting participants for clinical trials is governed by established ethical norms and protocols. Out of these, the moral right of people participating in a trial to make their own choices to participate is the most important. Informed consent is a critical aspect of the research ethic. The basic principle behind informed consent is to protect the autonomy of human subjects and, in particular, to ensure that the welfare and interest of the person are always put above

¹⁰ Based on Domecq et al. [334], at point 22.

Area of engagement	Platform
Identifying research priorities	James Lind Alliance (http://www.jla.nihr.ac.uk/) PCORI (www.pcori.org)
Designing and undertaking research	PatientsLikeMe (www.patientslikeme.com) 23andME (www.23andme.com) OMERACT (www.omeract.org) Reg4All (https://www.reg4all.org/) Sage Bridge Platform (http://sagebionetworks.org/platforms/)
Improving access to clinical trial data	European Union (www.clinicaltrialsregister.eu) Trials 4 Me (http://trials4me.lillycoi.com/) NIH (https://clinicaltrials.gov/)

Table 3.
 Selected online initiatives designed to engage patients in clinical trials.

society’s interest. In practical terms, that means society’s betterment can never be built on sacrificing the rights and health of research participants [337]. Strict internal procedures are in place to review all clinical trial proposals and to ensure that adequate informed consent procedures are followed. The ethical principles governing informed consent have important repercussions for data sharing.

The procedure for obtaining informed consent includes obtaining participants’ approval for data sharing and/or data archiving. It is a lengthy procedure and before enrolling, each prospective participant in a clinical trial must, among other things, be provided with a statement describing:

- The management of confidentiality of the collected information
- How records (data) that identify the participant will be kept
- The possibility that regulatory agencies may inspect the records.¹¹

Since potential future uses of clinical data cannot always be known in advance, the patients’ consent for data sharing can create a tension for ethical research practice.¹² At the same time, data archiving can also be seen as ethical practice, because it ensures that data is collected in line with rigorous methods and that the data is utilised to the maximum extent possible. This can help to develop new treatments and avoid unnecessary wasting of time and resources, including the time of research participants.

What is more, clinical research practitioners are well-versed in the potential risks resulting from the unauthorised sharing of data. Mechanisms have been developed to anonymise data so as to protect the privacy of research subjects and prevent data matching.¹³ The emphases on privacy controls and patient consent are unique to research using clinical trial data and pose challenges in other forms of data

¹¹ See, for example, the requirements of the US Federal Food and Drug Administration [338].

¹² The institutional review board or independent ethics committee reviews a research proposal to ensure that adequate informed consent procedures are determined and implemented in an ethical way without jeopardising the rights, safety, and well-being of the human subjects.

¹³ Data matching involves bringing together data from different sources, comparing it, and possibly combining it.

sharing. At this point, research ethics committees in the United States tend not to approve informed consent documents that require sharing of patient-level data. The reasons put forward for this practice are the ethical considerations for protecting participants and efforts to minimise any risks arising from data sharing ([339], p. 54).

In the past, patients participating in trials were required to sign broad consent forms, which left them little control over data and its future use. With the increasing focus on patient-centred health care, there has been a shift towards allowing patients to decide how their information and clinical data is used and to manage these permissions online. Initiatives such as Reg4All and Sage Bionetwork Bridge have developed digital forms for obtaining informed consent. These forms enable patients to decide who can use their data and how they use it. Reg4All also allows patients to track the uses of their data [340].

The early indications from these experiments are positive. Over 70% of patients sharing their data via Sage Bionetwork Bridge have opted to share data broadly and not to limit user access [341]. This finding also suggests that the universal adoption of electronic data sharing in clinical trials as envisaged by John Wilbanks, the founder of Sage Bionetwork, may indeed be feasible. His particular objective is to collect medical data from health-tracking and health-monitoring devices and applications installed on the smartphones of trial participants.

One limitation to this development is that the process of granting informed consent electronically is quite cumbersome, due to the considerable scope of informed consent. Sage Bionetwork enables the consent form to be processed on a mobile device, but that involves over 20 consent steps.¹⁴ The points the potential data providers are asked to consider with regard to data sharing include:

- What exactly is being shared?
- What purpose or purposes are being served by sharing?
- Who 'owns' the data? Can it be sold?
- Who is responsible for the security and privacy of the data?
- Where will the data be warehoused?
- Who will have access to the data?
- What happens if I change my mind?
- How am I protected if my data is disclosed?¹⁵

These new approaches to obtaining digital consent allow participants to choose whether to share data but also to participate in the clinical trials, irrespectively of their responses to the question on sharing.

However, in a clinical research setting, there is an ongoing debate surrounding the issue referred to as 'compound consent' [342].

The primary argument for *not allowing* patients to participate in the trial if they elect not to share their data is that this would lead to discrepancies between the

¹⁴ *Ibid.*

¹⁵ *Ibid.*

findings produced by the original research team and any subsequent uses of the data. Specifically, it would be impossible to conduct any secondary analyses of the dataset collected during the trial and to verify the original findings.

The primary argument for *allowing* patients to participate in a clinical trial if they elect not to share some or all of their data is to build trust among trial participants, especially those from vulnerable social groups or those with sensitive diseases. Participation by these people in clinical trials is crucial to the development of new medicines targeting such specific conditions. In these cases, it has been argued that compound consent might be an appropriate way to achieve greater enrolment in clinical trials.¹⁶

The conditions of informed consent may need to be adjusted to specific clinical trials and to specific patients group, allowing patients to have greater control over their data. This appears to be the latest best practice, but it poses some limitations on future data sharing and on the replication of outcomes achieved in earlier clinical trials.

6.2.2 Regulatory agencies

The ultimate objective of commercial clinical trials is to gain regulatory agency approval of pharmaceuticals for human use. Sponsors of clinical trials seeking regulatory approval from agencies such as the European Medicines Agency (EMA) or the U.S. Food and Drug Administration (FDA) must provide detailed clinical study reports (CSRs) with data on individual participants in their marketing applications for new products. Regulatory agencies are therefore in a strong position to influence the conduct of clinical trials, including data-sharing principles. Up until very recently, the practice of gaining marketing and regulatory approvals for pharmaceuticals has been well-defined and rigorous, yet it is not particularly transparent in terms of enabling public access to the materials submitted by companies as part of the approval process. The landscape has changed completely in recent years, with the EMA leading the way in facilitating open sharing of clinical trial data globally.

In the past, the agency was criticised for not releasing sufficient information after receiving external requests [343]. It is possible for any European Union citizen to request information from any European Union institution, including the EMA.¹⁷ Nonetheless, the EMA was not releasing a sufficient degree of information in response to such requests.

In early 2010, the European Ombudsman considered that the agency's repeated refusals to disclose public documents constituted acts of maladministration [344]. The agency's reasoning had been that the documents requested fell under the exceptions contained in the Rules for the Implementation of Regulation of the European Commission on access to EMA documents [345]. In its decisions to refuse access, the EMA had invoked Article 3(2)(a) of the rules, which refers to the protection of 'commercial interests of a natural or legal person, including intellectual property'.¹⁸

However, the Ombudsman held that the EMA reasoning was unconvincing, given that the study reports and protocols requested did not appear to involve any commercial interest. The Ombudsman ordered that the complainants be granted access to the clinical study reports and corresponding trial protocols.

¹⁶ Institute of Medicine [339], at point 29, 51.

¹⁷ See Article 255 of the treaty establishing the European Community.

¹⁸ See EMA/MB/203359/2006 Rev. 1 Adopted.

In response, the EMA later that year published a new policy governing access to the documents it holds [346]. This was a significant step towards greater transparency. In under 5 years since then, the EMA has released more than 1.9 million pages in response to such requests [343].

Another milestone was the announcement by the EMA Director in November 2012 to the effect that the agency was committed to broader data sharing in order to ‘rebuild trust and confidence in the whole system’.¹⁹ This was followed, in July 2013, with the release of a draft policy on the publication of and access to clinical trial data—defined as both clinical study reports and de-identified individual participant data—immediately after the regulatory decision by the EMA is made [347]. In the draft policy, the EMA stated that CSRs do not contain commercially confidential information and therefore could be released with no redactions.

In response to the draft policy, the EMA received over 150 submissions focusing on three areas—firstly, protection of patient privacy, secondly, whether information contained in CSRs could be considered commercially confidential and be used by competitors for commercial advantage, and, thirdly, the legality and enforceability of the data-sharing agreement between the EMA and data users [348].

Following extensive consultations, the final policy was released at the end of 2014 and came into effect on 1 January 2015, with the first reports made available in late 2016 [349]. The policy applies to clinical data—composed of clinical reports and individual patient data and statistical methods submitted under the centralised marketing authorisation procedure (covering the whole of the European Union).

Under the policy, the EMA will make available CSRs with commercially confidential information redacted, after making a regulatory decision on whether to grant or to refuse marketing authorisation.²⁰ Any member of the public may view this information on the EMA website or download it after registering and agreeing to the terms of use.

The key aspects of data use include:

- The user may not seek to reidentify the trial subjects or other individuals from the clinical reports in breach of applicable privacy laws.
- The clinical reports may not be used to support an application to obtain a marketing authorisation and any extensions or variations thereof for a product anywhere in the world.
- The clinical reports may be used solely for academic and non-commercial research purposes.²¹

Individual participant data is not yet available online. The EMA plans to make this available in late 2019–2020, depending on the completion and testing of an online portal that needs to be consistent with the European Union regulations protecting individual privacy.

The recent EMA policy has already faced multiple legal challenges from pharmaceutical companies. The issue of commercial confidentiality was also recently adjudicated in the General Court of the European Union. The court recently delivered three well-considered judgements in cases brought by companies objecting to the disclosure of their documents and data submitted to the EMA. The court

¹⁹ *Ibid*, 8.

²⁰ *Ibid*.

²¹ *Ibid*, Annex 2.

dismissed all three cases and considered the companies had failed to provide any concrete evidence of how the disclosure of the contested documents would undermine their commercial interests. Those cases are analysed in Section 7.3.

The developments in Europe have stimulated expert discussion around data sharing in other parts of the world. In the United States, the Institute of Medicine released a report on the sharing of clinical trial data, which is the most authoritative study on the subject to date. The study was released in 2015, following extensive consultations.²² Nevertheless, the FDA does not currently support the open sharing of clinical reports or data submitted as per the marketing approval process, referring to the strict trade secret and personal data protection laws in the United States.

Data submitted to the FDA is governed by two statutes—the *Freedom of Information Act*²³ that deals with disclosure in response to citizen requests and the *Trade Secrets Act*²⁴ that limits affirmative disclosure by the government. A relevant regulation defining the constraints is 21 CFR 20.61(c), which provides that:

... data and information submitted or divulged to the Food and Drug Administration which fall within the definitions of a trade secret or confidential commercial or financial information are not available for public disclosure [350].

Moreover, the authors of the 2015 study concluded that these statutes, together with case law interpreting them, generally prohibit regulatory agencies such as the FDA from releasing information that is likely to cause substantial harm to the competitive position of the person from whom the information was obtained. However, unlike EMA, the FDA has not yet tested whether the release of CSRs and other data provided as part of the approval process may be eligible for release since it rarely includes patient-level data.

One important issue raised in the 2015 report is whether the FDA has the authority to issue regulations that could override the constraints of the *Trade Secrets Act*, which does allow federal government agencies in the United States to disclose trade secrets and confidential information unless such disclosure is authorised by the law.²⁵

Scholars have put forward two scenarios under which the FDA might be able to do so. Firstly, the FDA potentially has the authority to disclose trade secrets for public health reasons, resulting from the provision in the *Hatch-Waxman Act* stating that the FDA is supposed to release clinical trial data after the data exclusivity period expires [351]. Secondly, the FDA has expansive authority to impose on regulated parties ‘other conditions’ that ‘relate to the protection of public health’.²⁶

In the past, the FDA has relied on this authority to propose disclosure rules for clinical data concerning human gene therapy.²⁷ In 2001, the FDA argued that several significant public health goals would be served through greater disclosure of data.²⁸ However, following publication of these analyses, the FDA withdrew the

²² Institute of Medicine [339], at point 29.

²³ 5 USC § 552.

²⁴ 18 USC § 1905 (1982).

²⁵ 18 USC § 1905 (1982).

²⁶ Under Section 501(i) of the *Food, Drug and Cosmetic Act*.

²⁷ See FDA ‘Proposed Rule on Availability for Public Disclosure and Submission to FDA for Public Disclosure of Certain Data and Information Related to Human Gene Therapy or Xenotransplantation’, 66 Fed. Reg. 4692 (2001).

²⁸ *Ibid.*

Rule on Availability for Public Disclosure that it had issued in 2001 when making the human gene therapy decision on data publication [352]. The decision thus closed the second possible option for releasing clinical trial data submitted as part of the FDA approval process.

In early 2018, the FDA announced plans to conduct a series of pilots to release portions of CSRs redacted by the FDA to exclude confidential commercial information, trade secrets, and personal data that can fall subject to privacy laws [353]. Patient-level data is not included in the pilots.

One final comment on the role of regulatory agencies in setting the standards for data sharing is that only a small proportion of all clinical trial data is submitted to regulatory authorities. Most academic and publicly funded trials are not designed with the objective to seek regulatory approvals. Therefore, the data presented to regulatory agencies represents only a small subset of the clinical trial data available globally.

At the same time, the data is of high commercial value and therefore has the significant potential to influence future practice and study—which highlights the importance of regulatory agencies in the open data debate.

6.2.3 Industry partners

Another important stakeholder group consists of the industry partners who spent significant resources on clinical trials with the sole objective to develop and market new products. For these reasons, the information produced in clinical trials is regarded as confidential, and the culture of data sharing is not favoured by industry.

The reasons this group puts forward for confidentiality include:

- The documents from clinical trials may hold considerations based on confidential interactions with regulatory authorities.²⁹
- The documents may contain data that is subject to personal protection and informed consent.
- The data may disclose internal business or internal scientific expertise and business development strategies.
- Access to data might lead to conflicting or incorrect secondary uses by non-qualified users.
- Participating researchers and investigators wish to be able use the data in articles about the trials, which is important to career progression and is an important incentive for researchers to participate in industry-led clinical trials.

A further factor put forward in the argument against data disclosure was that industry partners like to use previous clinical data in the development of subsequent products and access to such information could give competitors an advantage—thus shortening the lead time between the marketing of the first product and when similar products begin to appear based on the original.³⁰

²⁹ Institute of Health at point 29, 62.

³⁰ *Ibid.*

Scholars and industry partners alike have argued that the lead time from ‘research to market’ is the key factor facilitating return on investment in research and development [354]. That lead time has already decreased substantially in recent years and further decreases through data disclosures might discourage further investments in drug development.³¹ This issue is of particular importance to the development and marketing of biosimilars and is further discussed in Section 7.3.

To date, most data sharing by pharmaceutical companies is based on internal policies that require researchers to submit detailed proposals, which are subject to strict internal reviews by the company and often limit the release to highly redacted data. In some cases, external investigators have obtained access to industry-sponsored clinical trials and have conducted independent secondary analyses that identified significant issues in the underreporting of negative results and serious side effects.

These studies also have reported industry failure to publish the results of negative trials for widely prescribed therapies [355]. While industry groups have denied such claims, the subsequent sparring has, in several cases, initiated further clinical trials to test the contested issues and led to billion-dollar settlements of legal disputes.³² In some other cases, changes in the labelling of pharmaceuticals or restrictions to prescribing certain drugs to risk-prone patients were required.³³

Such grave instances of professional misconduct and manipulation of clinical trial data further support the case for releasing clinical trial data.

It the meantime, complaints from researchers and investigators on repeated denial of access to clinical trial data have initiated discussions among industry partners to introduce more transparent approaches to data release and expert sharing.³⁴ The open-sharing policies released by the EMA also played a major role in changing industry approaches to data sharing. Major pharmaceutical industry associations in Europe and the United States have committed to the sharing of clinical trial data and have agreed to develop a process for data sharing [357].

One issue in urgent need of addressing is the cost of sharing clinical trial data. There are several cost components to open clinical data, and additional human resources are required to redact confidential information and personal data from reports and data.

Other costs are associated with preparing and reviewing the reports and data for publication—including due diligence and payments to external reviewers and auditing panels to ensure compliance. Additional work is required to prepare templates for informed consent that would enable greater data sharing and to provide lay summaries of the data.³⁵ Almost half of the applications submitted for FDA approval are filed by small businesses [358], and the submitters may not have the resources for preparing clinical trial data for publication.

6.2.4 Other stakeholders

There are many other stakeholders with the leverage to set standards and to encourage the sharing of data arising from clinical trials. In addition to those

³¹ *Ibid.*

³² See Table 3.1 ‘Examples of Effects of Independent Analyses Carried Out on Clinical Trial Data’, in Institutes of Health (2015) at point 29.

³³ *Ibid.*

³⁴ For a good summary, see Krumholz et al. [356].

³⁵ The costs detailed by industry are itemised in Institute of Medicine [339] at point 29, 68.

mentioned already, research funders and publishers are significant stakeholders. Their role in data publication and release is outlined in Chapter 3.³⁶

Other important stakeholders include researchers and investigators involved in clinical trial design, collection, processing, and analysis of data. Research investigators are not involved in the data collection but typically use it for secondary analyses and reanalysis of the results, which are typically published in scientific publications. The work of clinical researchers and investigators is largely directed by industry partners. Again, this raises the issue of funding for data sharing. In the 2015 study, researchers raised grave concerns that if data sharing becomes an unfunded mandate, the costs of sharing will reduce the funding available for new grants. This would, in turn, result in fewer new trials.³⁷ The study further noted that this concern is particularly cogent for researchers working in low-resource settings such as those affected by neglected global diseases.³⁸

Medical research institutes and universities also have an important stake in the sharing of clinical trial data. They can influence data sharing through infrastructure support, providing incentives and training to researchers and by conducting scientific reviews. Universities employ many investigators involved in clinical trials, and they often provide the infrastructure for the design and execution of clinical trials.

Public research organisations and universities currently provide relatively little support for data curation, documentation, and sharing. While they may have created a research data management function in libraries, these are not, at this point, staffed by expert data analysts who could assist researchers with the documentation and preparation of the data for publication. What is more, there is currently little or no training provided to researchers in the procedures, documentation methods, and structures needed to share data. Even if such training modules were available, they would have to compete for the researchers' time. Given that the incentives for data sharing are minimal, data preparation is generally not a priority for researchers. These issues are examined in the section dealing with incentives below.

Other stakeholder clusters with an increasing role in clinical trials are the patient advocacy groups and associations dealing with rare diseases. These are not-for-profit organisations, foundations, or even loose networks that aim to raise funds for research and study and to provide patient education and support for clinical care. Patient advocacy groups are active in the recruitment of patients for clinical trials. In recent years, these groups have become more active, such as creating online platforms for patient engagement (see **Table 3**). In the United States, patient advocacy groups have become more active providers of the data collected directly by their members—such as from their smartphones or medical measurement devices [359]. With increasing demands for participatory medicine, patient advocacy groups will likely play a more important role in clinical trials into the future.

6.3 The stages of data sharing

From the discussion above, it is apparent that the sharing of clinical trial data occurs through various networks and platforms—from sharing data underpinning publications, through providing extensive clinical summary reports and individual data to regulatory agencies as part of the marketing approval process, to sharing clinical data directly between industry partners and research investigators and to depositing data online in discipline or research-specific repositories.

³⁶ Especially Sections 3.2 and 3.4.

³⁷ Institute of Health at point 29, 72.

³⁸ *Ibid*, 73.

Similar to the stages of processing and release of particle physics data, described in the previous chapter, clinical trial data is also being shared along different stages of processing, granularity, and control. Such data can also be shared openly, by depositing it online or upon request. Given these similarities, and for the purposes of consistency, the sharing of clinical trial data is described below along four levels, with Levels 1, 3, and 4 roughly corresponding to data-sharing levels at CERN. Level 1 is the data underpinning publications, Level 3 consists of analysable datasets that enable full reproducibility of analyses, and Level 4 represents raw data (or experimental data) collected in the course of research studies.

6.3.1 Data Level 1: data underpinning publications

Given that changes to policies on releasing clinical trial data by regulatory agencies are very recent and are not yet fully implemented in practice, publication in peer-reviewed scientific journals remains the primary method for sharing clinical trial data with the scientific and medical communities, as well as with the public. Typically, several publications are written in the course of a clinical trial.

The first publication usually covers the key objectives of the trial and reports the key outcomes and baseline measures. Subsequent publications focus on a specific aspect of the primary analyses and report outcomes for a particular subgroup of patients.³⁹

The deadline for the release of data underpinning publications varies among publishers and research funders. The study by the Institute of Medicine recommended depositing data within 6 months from publication.⁴⁰ The WHO Joint Statement on public disclosure of results from clinical trials includes an ‘indicative timeframe’ of 24 months from study completion, to allow for peer review [360]. Funders and publishers generally tend not to include a specific deadline in their policies, as discussed in Chapter 3.⁴¹ Instead, research funders tend to use wording such as ‘within a reasonable timeframe’.⁴² The reason for this is that researchers have argued that they require time to write publications resulting from their research and that their careers may be jeopardised if others use their data before the publications are finalised. This study summarises recommendations in this regard in Sections 8.3.4 and 8.3.5.

Differences also exist among funders, publishers, and researchers as to what they consider ‘data underpinning publications’. While funders generally tend to suggest that ‘data’ refers to any corresponding dataset, which includes data supporting the claims made in the publication, the Institute of Medicine has further clarified the meaning of Level 1 data. Its study refers to this data as the ‘post-publication package’—which, in the view of the review committee, should consist of the ‘analytic dataset and metadata, including the protocol, statistical analysis plan,⁴³ and analytic code, supporting published results’.⁴⁴

³⁹ Institute of Medicine at point 29.

⁴⁰ *Ibid.*

⁴¹ See Section 3.4.

⁴² National Science Foundation (2010).

⁴³ Statistical analysis plan describes the analyses to be conducted and the statistical methods to be used in a clinical trial. It typically includes plans for analysis of baseline descriptive data and adherence to the intervention, prespecified primary and secondary outcomes, and definitions of adverse and serious events and comparison of these outcomes across interventions for prespecified subgroups.

⁴⁴ Institute of Medicine at point 29, 108.

However, it is unclear whether the recommended post-publication dataset should include everything required to reproduce the published results, as published, or whether it should simply include enough evidence to support the findings. The committee further shared the view of the National Research Council, which in 2003 stated that:

Community standards for sharing publication-related data and materials should flow from the general principle that the publication of scientific information is intended to move science forward. More specifically, the act of publishing is a quid pro quo in which authors receive credit and acknowledgment in exchange for disclosure of their scientific findings. An author's obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research [361].

As such, the review committee stated that publication data should ideally be shared immediately after publication.

6.3.2 Data Level 2: summary data

Level 2 data in clinical trials is represented data, which can include any of the following—lay summaries of results to be published in prescribed registries (such as ClinicalTrials.gov), clinical study reports, either full or extracts from these reports, or summaries of clinical trials aimed at the general public.

All clinical trials are subject to requirements to report the results in registries, usually multiple registries, depending on the sources of funding and the approvals sought following completion of the trials. These summary reports are typically available on the registry website and generally are limited to major outcomes and adverse events.⁴⁵

The approaches taken to the publication of summary reports differ sharply between Europe and the United States.

In the United States, the *Food and Drug Administration Amendments Act* (FDAAA) requires results of trials of FDA-regulated products to be reported to ClinicalTrials.gov within 12 months of study completion.⁴⁶ In 2016, the policy was extended to apply to *all* clinical trials funded in whole or in part by the National Institutes of Health—regardless of study phase, type of intervention, or whether the trial is covered under the FDAAA [362]. The reasoning put forward for such extended reporting was that:

... when research involves human volunteers who agree to participate in clinical trials to test new drugs, devices, or other interventions, this principle of data sharing properly assumes the role of an ethical mandate [363].

The data elements to be provided under the legislation include participant flow, demographic and baseline characteristics, outcomes and statistical analyses,

⁴⁵ Adverse events are 'unfavourable changes in health, including abnormal laboratory findings that occur in trial participants during the clinical trial or within a specified period following the trial' (ClinicalTrials.gov, 2014). See also Institute of Medicine [339], p. 107.

⁴⁶ Under Section 402(j) of the *Public Health Service Act*, as amended by Title VIII of the *Food and Drug Administration (FDA) Amendments Act* of 2007 (FDAAA), and the regulation Clinical Trial Registration and Results Information Submission, at 42 CFR part 11.

adverse events, the protocol and statistical analysis plan, and administrative information.

The revised mandate accommodates delays of data submissions for up to an additional 2 years for trials of unapproved products for which initial FDA marketing approval or clearance is being sought.⁴⁷ Despite the fact that the law provides for hefty fines for missing data submissions, compliance with the prescribed timeframe was lagging.

In Europe, the release of summary data is driven by the EMA. Under the 2015 EMA data sharing policy outlined earlier in this chapter, the EMA is fully committed to publishing clinical reports submitted to the agency as part of the marketing approval process. The EMA defines clinical reports as comprising these documents:

Clinical overview—a critical analysis of the clinical data submitted. This should present the risks and limitations of the medicine development programme and the study results, analyse the benefits and risks of the medicinal product in its intended use, and describe how the study results support critical parts of the prescribing information.

Clinical summary—a detailed factual summary of the clinical information. This should include information provided in clinical study reports from any meta-analyses or other cross-study analyses for which full reports have been provided in the submission, as well as post-marketing data for products that have been marketed outside of the European Union.

Clinical study report—a detailed document about the method and results of a clinical trial. It will be a scientific document addressing safety and efficacy and its content. This should include several appendices, of which three should be published online, as follows:

1. **Protocol and protocol amendments**, describing the objectives, design, methods, statistical considerations and organisation of a clinical trial.
2. **Sample case report form**, a questionnaire used by the sponsor of the clinical trial to collect data from each participating site.
3. **Documentation of statistical methods**, providing a description of the methods used for collection, analysis, interpretation, presentation, and organisation of the data [364].

The EMA is the first regulatory agency to require full publication of CSRs, and given the breadth of the information required for online publication, the data is likely to be meaningful to diverse audiences for further study and analysis. The required data is to be published within 60 days after the European Commission decision on marketing authorisation or within 150 days after the receipt of the withdrawal letter, as stated in **Table 4**.⁴⁸

The summary data under the EMA data-sharing policy is currently published on the EMA clinical data portal [365], which will be, in the near future, substituted for a new portal to enable publication of individual patient data (Level 4 data) in accordance with strict privacy laws in the European Union. The new portal is expected to be operational in 2019.

⁴⁷ NIH at point 86.

⁴⁸ *Ibid.*

	Data type	Data limitations	Data availability
level 1	Data underpinning journal publications (underlying data) typically report the main outcomes of clinical trials.	Lack of detailed data and information about study design, efficacy and safety analysis.	Based on individual policies of research funders and publishers. Compliance with the policies is currently lagging.
level 2	Summary data EMA: Clinical reports including clinical overviews, clinical summaries, and clinical study reports (CSR), together with appendixes to the CSRs. FDA: Lay-language trial summary FDA: PILOT: Portions from CSR that contain complete summaries of the study results, the protocol and protocol amendments, and the statistical plan.	Data redacted to exclude commercial confidential information detailed in Annex 3 of the EMA policy. Only high-level results, without supporting data PILOT: Data redacted by FDA to exclude confidential commercial information, trade secrets, and personal information.	EMA: 60 days after the EC decision on marketing authorisation or within 150 days after receiving the opinion of the Committee for Medicinal Products for Human Use. FDA: High-level summaries to posted on www.clinicaltrials.gov within 12 months of study completion. No mandate to release CSR at this stage.
level 3	Analysable datasets – pre-processed data prepared to address specific research questions. Include individual patient data. Related software and algorithms need to form part of the dataset.	Resource intensive: require detailed documentation of all the steps and assumptions taken to collect and process the data, including statistical methods, algorithms, tools and protocols. Require protection of privacy of research subjects.	EMA: expects to publish patient level data from 2019/2020. Questionable whether covered by open data mandates of research funders and publishers.
level 4	Raw data, including individual patient data	Full potential for research reproducibility. Require protection of privacy of research subjects.	Not routinely shared with the broader scientific community or the public, but portions of the data can be released under the Freedom of Information legislation.

Table 4.
Levels of access to clinical trial data.

6.3.3 Data Level 3: analysable datasets

Level 3 data refers to pre-processed data prepared to address specific research questions. Such data may be a subset of the original dataset collected in clinical trials, or it can be a full dataset after application of defined cleaning and processing steps and after performing statistical analyses. Related software, metadata, and algorithms need to form part of the dataset.

Level 3 data, or subsets thereof, are required to reproduce the results published in CSRs or in publications. However, research organisations, in general, and industry partners, in particular, are generally reluctant to share Level 3 data with external parties.⁴⁹ This data represents the full potential for secondary data analyses, and the risks associated with data sharing and reuse are greater than those with lower level data.

Of particular importance to regulatory agencies are the tasks of developing mechanisms to mitigate the risks associated with the protection of the privacy of research subjects and to avoid potential errors of misinterpretation of secondary

⁴⁹ The reasons put forward by industry for not sharing data are discussed in Section 6.3.3 of this book. The reasons put forward by researchers for not sharing data are discussed in Section 6.4.4.

analyses. The EMA is the most advanced in developing such mechanisms and expects to publicly release patient-level data in late 2019.

At this point, however, the sharing of Level 3 data is limited to expert users and is usually delayed for several years after the initial study date, especially if secondary trials are still under way. For these reasons, the policy developed by the EMA may present a momentum in driving the release of reusable clinical data globally.

6.3.4 Level 4 data: raw data, including individual patient records

The lowest level data in clinical trials is the data collected directly from patients or obtained through other means, such as by sourcing the data from medical equipment or from health applications of patients. This data may or not be cleaned, preprocessed, or packaged. It is, however, usually structured, due to the prescribed design of clinical trials. Access to raw-level data and patient records would be highly beneficial for designing new clinical trials based on previous data. While Level 3 data can also offer such functionality, the reuse of Level 3 data may be limited to the cleaning and statistical techniques taken by the original research team. By contrast, Level 4 data enables the design of new data cleaning and processing protocols.

Level 4 data is generally not shared with external audiences, except in expert settings. Portions of Level 3 and/or Level 4 data may be obtained under freedom of information legislation. In the United States and Europe, the regulatory agencies generally disclose non-summary safety and efficacy data from a specific application only in response to such requests. The potential for sharing this type of data as open data will increase when experiences with sharing and reusing higher-level data emerge. Of particular interest will be efforts to allow the reusability of Level 3 data while protecting the privacy of research subjects.

The issues of privacy are covered in detail in Chapter 7, Section 7.5, of this study. Other challenges associated with the sharing of clinical trial data are discussed below.

6.4 The challenges of open sharing of clinical trial data

6.4.1 Demands on researchers and changing attitudes towards data sharing

Curating and sharing research data, in general, and clinical trial data, in particular, pose several challenges to researchers. Many of these are interconnected with the challenges of data management described above.

The key challenge is that preparing and maintaining usable data repositories require a great deal of effort and resources, especially the time of researchers who collected the data and need to describe it in meaningful ways to make the data legible to prospective users while ensuring simultaneous compliance with the open data mandates and any confidentiality, privacy, and internal policies that may apply. This is a time- and labour-intensive process, especially in organisations implementing controlled access to data. Receiving and processing applications for data release, developing agreements and contracts, producing and transferring data, and responding to subsequent requests for clarification involve a broad range of people across the data-sharing organisation [366]. In this sense, open data more resembles new data products than readily available outputs from previous experiments. Research organisations typically have limited resources to handle these requests, which can result in conflict with other demands on staff time, such as ongoing research. In the absence of support from research funders to prepare the datasets, some research units may require applicants to provide funding to cover

the necessary staff time, which in effect means spending research money on curation.⁵⁰ In the absence of funding specifically to curate and release data as open data, researchers continue to share their data with others in different ways.

Recent surveys of researchers across many disciplines also confirm this finding. For example, in a 2016 survey conducted by Wiley, over 4600 researchers reported the three most common ways of sharing data as via conference (the box ticked by 48% of respondents), as supplementary material in a journal (40%), or, informally, upon request received from other colleagues (33%). Only 20% of researchers ticked the box stating they had shared data formally via an open access repository, whether via institutional, discipline-specific, or journal repositories.⁵¹ The authors of the study concluded that:

*... these results demonstrate that researchers continue to be unclear on what [open] 'sharing' data means in the sense of providing unlimited, appropriately licenced and permanent access to their data.*⁵²

At the same time, data sharing among researchers has markedly increased in recent years. The same survey found that 69% of researchers said they had shared data from their research in some way. This represents an increase of 17% from the same survey conducted by Wiley 2 years earlier, in 2014.⁵³ A similar result was reported in a survey of over 1200 researchers by Elsevier and the Centre for Science and Technology Studies. Specifically, 65% of respondents said they had previously shared their data with others [369]. At the same time, over one-third of researchers did not share their data from their last projects.⁵⁴ The attitudes of researchers to data sharing are summarised in **Figure 8**.⁵⁵

Another interesting finding of the two studies is the reasons why researchers prefer to share (or not to share) their data. Both surveys confirm that researchers share data because it is an established practice in their field of science, such as genomic research, or that the mandates of publishers or funders require data sharing. Other primary reasons for data sharing include an ethical and moral responsibility to share data and the public benefits likely to accrue from data sharing. The Elsevier survey also established that when researchers share their data directly, the vast majority—over 80% of researchers—choose to share data with direct collaborators. This suggests that trust is an important aspect of sharing data and that credit and increased visibility are not major motivators for data sharing, as discussed below.

6.4.2 Incentives for researchers

The greatest challenge to data sharing is, arguably, the lack of incentives for researchers to share data. What is even worse, many researchers do not see value in data sharing, as documented by survey results and many authoritative studies. For example, a recent survey of researchers who received grants from the Wellcome Trust showed that the potential loss of publication opportunities, along with the

⁵⁰ *Ibid.*

⁵¹ The results of the survey are reported here: Wiley Open Science Researcher Survey [367]; and Wiley Network [368].

⁵² *Ibid.*

⁵³ *Ibid.*

⁵⁴ *Ibid.*, 20.

⁵⁵ Source: Elsevier and the Centre for Science and Technology Studies [369] at point 98.

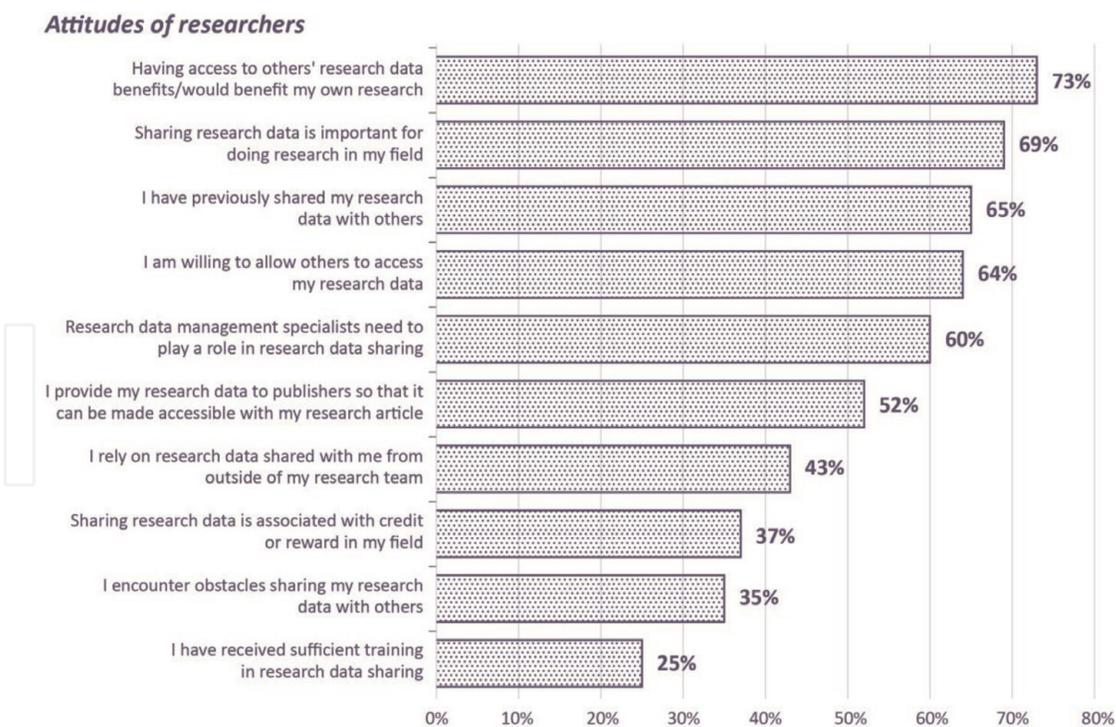


Figure 8.
Attitudes of researchers to data sharing (measured as % of respondents who agree with the statement).

belief that publishing is the only criterion for successful grant funding and academic advancement, was the major factor in the inhibition of data sharing [370].

Much effort has gone into the development of data citation practices and into measuring impact through data citation [371–376]. Yet researchers have little reason to value data metrics (including citations), because increased data impact is not the incentive for data collection and sharing in the first place. Rather, data collection is necessary to conduct research and write publications for which research grants are provided, for which researchers are rewarded, and upon which the careers of researchers depend. As noted by the Institute of Medicine [339]:

In the eyes of performance review and promotion committees, the primary criteria for academic success rest on publications, funding, leadership, and teaching. Data sharing is not an activity that receives attention from promotion committees, and there is insufficient recognition of the intellectual effort involved in designing, accruing, curating, and completing a clinical trial data set. In this way, the lack of incentives for sharing clinical trial data is analogous to the recognised dearth of incentives for team science within university settings.⁵⁶

Furthermore, researchers need reasonable time and exclusivity to work with the data they collect and to achieve the outcomes of their original research before they can share that data with others.

For these reasons, data and impact citation practices are unlikely to promote the curation and release of open scientific data and to promote collaboration among researchers, as further explained in Chapter 8, Section 8.3. Subsequently, this study argues that acknowledging the original data collectors as co-authors of any subsequent publications arising from the data reuse might be a more effective way to promote the release of research data as open data.

⁵⁶ Institute of Medicine at point 29, 76.

6.4.3 The limits of research reproducibility

Achieving research reproducibility is one of the objectives put forward in support of open scientific data.⁵⁷ In clinical trials there is a huge interest in replicating results, initiated by some contested evidence that the results of clinical trials for new medicines presented in marketing approvals may be wrong, or at least misleading, as discussed earlier. If a scientific result can be confirmed or disproved by sharing the underlying data, then disputes could be resolved faster, some have argued [377].

However, efforts to reproduce results often reveal research and data processing disagreements on which these disputes are based. Interrogating data may not be the answer to this. In fact, the early experiences with data reproducibility suggest that this may not be the golden key in verifying research results, despite the fact that some researchers have proclaimed reproducibility as the gold standard for future science [378, 379].

The nub of the problem with research reproducibility is the lack of agreement across scientific disciplines on how to document research data so as to ensure its reuse by independent users. The parameters for data reuse are far broader than just sharing metadata and supporting documentation, as discussed further in Section 8.3 of this book.

Other experts in the field, such as Borgman, have pointed out the problematic nature of reproducibility noting that very fine distinctions are made between validation, utility, replication, repeatability, and reproducibility, with each of these terms having a distinct meaning within individual scientific disciplines ([167], p. 209). The sharing of data resulting from clinical trials and genomic research is particularly challenging, due to a very large amount of data analysis that lead to meaningful discoveries [380]. On some occasions, the objective might be to replicate the original research undertaken, using the same techniques, while other approaches may aim to achieve comparable results using similar inputs and methods. The first approach would verify the published results, while the latter would confirm the hypotheses being tested, as Borgman also observed.⁵⁸

Clinical trials are very much concerned with replication, yet the necessary resources and the costs of reproducing research trials might be prohibitive. Raw-level data is typically required to achieve reproducibility, and such data is expensive to document and curate and may not even be available for sharing. The Institute of Medicine reported that biomedical companies often attempt to replicate the results reported in a journal article as a first step in determining whether a line of research is likely to be productive.⁵⁹ Such companies may spend millions of dollars, but the investments may well not yield any success.⁶⁰

In some ways, the calls for research reproducibility can also be seen as ‘reinventing the wheel’ rather than focusing on future research. By challenging the authority of previous research, trust among researchers that is the basis for sharing data may be undermined. In many circumstances, reproducibility has severe limitations and should not be the stated objective for open scientific data.

⁵⁷ See Chapter 2, Section 2.3.4.

⁵⁸ Borgman at point 106.

⁵⁹ Institute of Medicine at point 29.

⁶⁰ *Ibid.*

6.4.4 Managing ethical uses of open data

One of the risks raised by researchers as an impediment to data sharing is the possibility of data misuse or even wilful misinterpretation of the data. In the situation where there is a lack of agreement on what constitutes data reuse and reusability, these concerns are justified. In the absence of robust data descriptors, the data may be analysed incorrectly, and incorrect conclusions may be drawn. Another concern is that the purpose for which data is later reused may be incompatible with the original purpose for which the data was collected. Such purposes may include causes with which the original data creator disagrees or does not wish to be aligned. Researchers also raised concerns that future data users may not give proper credit to the original creators. A further concern is that data may be used to harm future business activities of research organisations, such as allowing others to commercially exploit the data.

At present, research organisations employ various methods to manage the risks. One approach for managing forms of permitted use can be the selection of appropriate licencing mechanisms, as discussed in Chapter 7, Section 7.3. While the Creative Commons Zero Waiver places a few limits on how the data can be reused, the broader suite of the Creative Commons 4.0 licences allows organisations to limit commercial uses of the data or the creation of derivative works. While such licences may limit the definition of open data, the sharing of research data for limited (non-commercial) purposes is still a better option than not sharing it.

Another alternative, recommended by the Open Data Institute in the United Kingdom [381], is to share data under CC-BY-SA 4.0 licence, which requires that any derivative works created from the research data be published under the same licence. This approach may help research organisations to manage undesired reuses of the data—for example, to use the data in commercial reports and to do so without attributing the original data creator.

In the field of clinical trials, data use agreements (sometimes also called data-sharing agreements) are often used to guard against the risks of data misuse. While these agreements go against the free sharing of open data, they do allow sharing for limited purposes. The key terms of such agreements in clinical trials include:

- Prohibition of any attempts to reidentify or to contact research subjects.
- No further sharing of the data.
- No commercial use.
- Requirements to attribute the original authors.
- No permission to brand any subsequent works as originating from the data producer.
- Non-endorsement of any future uses of the data by its originator.
- Secure handling and processing of the data.
- Assignment of intellectual property rights for discoveries made from the shared data.
- Limited warranties—no guarantees that data is fit for secondary uses, so reusers cannot claim damages if data is misapplied.

The last, and most effective, tool for managing the risks of data misuses and misinterpretation is through professional and ethical conducts supported by mandatory training. The complaints about data misuse can be resolved by appropriate disciplinary committees. After all, data sharing and reuse are a collective responsibility, and researchers have a key role in ensuring that data is reused ethically and in accordance with established norms and protocols.

Conclusion

Open data policies have renewed the focus on the sharing of clinical trial data, especially following the release of the EMA open data policy in 2014, with deep implications for clinical practice and research. The sharing of clinical trial data is now a more established practice within the discipline, as recent surveys confirm. The stages of data sharing and the responsibilities when sharing are clearly established across the entire research discipline, and there is a high degree of similarity in the data-sharing levels and practices among both public and private sector organisations.

At the same time, the sharing of research data in publications and directly with peers is still the method preferred over depositing clinical trial data in public repositories, especially at the level of patient data. The reasons for limited open data sharing identified in this chapter include:

- i. Safeguarding the privacy of research subjects
- ii. Ensuring compliance with confidentiality requirements of private research funders
- iii. Fear of potential unethical use and even the wilful misuse of the data by others
- iv. Lack of incentives for researchers to curate and share data
- v. Lack of funding allocated to open data curation and release
- vi. Lack of compliance with the open data mandates

The legal and privacy issues of data sharing are discussed further in the next chapter (Chapter 7), while Chapter 8 offers further insights into the issue of misinterpreted incentives and provides recommendations to address those issues.

Another finding of this chapter is that sharing clinical trial data does not necessarily lead to research reproducibility, as is often assumed by policymakers. Only data reusability can facilitate research reproducibility. However, low-level data is typically necessary to achieve reproducibility. Such data may not be readily available for sharing as open data or can be costly to curate. Therefore, reproducibility should only be the desired and stated objective in carefully selected research areas.

This chapter has further argued that the research profession has in place rigorous procedures for managing the privacy and confidentiality issues arising in the sharing of clinical trial data. Soft mechanisms such as codes of ethics and professional and research conduct, combined with data use agreements, have been found to be highly effective tools for managing the possible risks associated with the sharing of patient-level data. The new portal currently being tested by the EMA will showcase the world's best practice in enabling the safe and secure sharing of patient-level data as open data.

The successful sharing of clinical data as open data requires a combination of both approaches—professional and ethical data use, accompanied by robust technologies. Researchers are therefore the best positioned to control data sharing into the future.

IntechOpen

IntechOpen

Author details

Vera J. Lipton
Zvi Meitar Institute for Legal Implications of Emerging Technologies,
Harry Radzyner Law School, IDC Herzliya, Israel

*Address all correspondence to: vera.lipton@bigpond.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 