Clinical Trial Data Sharing Ecosystem

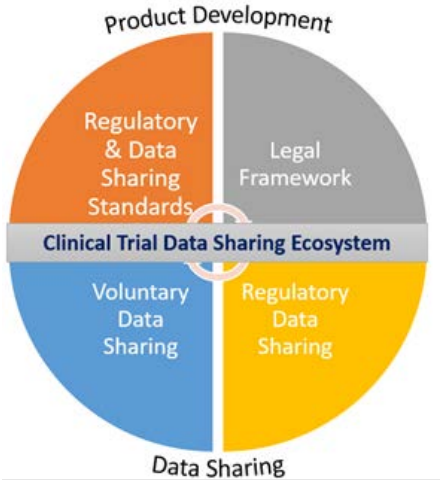

Table of Contents

1 Introduction

Years of criticism of lack of transparency about results of clinical research have been addressed by multiple requirements in terms of policies, legislations, regulatory guidelines, and multiple data sharing platforms for public disclosure. These platforms provide access to information about clinical research, from the registration of new and ongoing trials to the completed trials and sometimes the inclusion of result summaries, result reports and individual participant data upon request.

This paper aims to generate awareness of the established pharmaceutical industry data sharing ecosystem to include both voluntary and regulatory data sharing activities as such where a researcher can find information within that ecosystem. It also provides an overview of the legal framework of what information is in scope for each of the various platforms, that all serves compliance with policy or regulatory requirement.

2 Regulatory Standards

Regulatory authorities around the world have instituted regulations, guidance documents and consensus standards that govern what data should be shared and how it is exchanged in order for them to approve, monitor and regulate new and existing products (e.g., human drugs, biologics, medical devices). A key influence for health authorities' human drug or biological regulations for clinical research is the technical requirements for pharmaceuticals for human use issued by the International Council for Harmonisation (ICH). While most regulatory authorities closely follow the ICH guidelines for legislation, there may be differences with respect to final guidance to which the marketing authorization holders need to adhere.

ICH guidance includes many areas of a product's development, such as clinical and nonclinical research as well as manufacturing quality. The Common Technical Document (CTD) Triangle is a pictorial of the information needed to build a marketing application dossier that is submitted to regulatory authorities and became mandatory in some countries in 2003 [1]. The CTD triangle concept outlines various modules that build a foundation from individual study reports that support pooled summaries of information, which are then critically analyzed in an overview of the information. The focus of this paper is the modules of the CTD for clinical research and the ICH guidance that outlines the content of each module. See Appendix 1 for a representation of the clinical hierarchy with links to the associated ICH guidance for each CTD module.

2.1 End of Trial

At the end of a trial (EoT)/study completion, there are two major components that are the basis for the reports that are included in Module 5 of the marketing application dossier: the raw data from the trial (datasets generally not submitted to health authorities) and the clinical study report (CSR). Datasets are mapped and standardized according to Clinical Data Interchange Standards Consortium (cdisc.org) guidance in order to perform the predefined analyses in the statistical analysis plan (SAP)[2].

Key components of the CSR (per ICH E3 guidelines) [3] that are helpful to understand the trial design and summarization of the results are:

- CSR Synopsis – section 2
- CSR body – sections 3-15
- Appendix 16.1.1 - protocol (written per ICH E6 [R2]) [3]
- Appendix 16.1.2 – blank case report form (contains the form for the data that was collected and how it was mapped to standard datasets – per cdisc.org)
- Appendix 16.1.9 - statistical analysis plan (SAP) (written per ICH E9) [3]

The data from clinical trials are then used for marketing authorizations and mandatory disclosures and are often available to researchers for additional analyses via voluntary data sharing policies.

2.2 Data Sharing Standards

Public disclosure of clinical trial information and documents is required in countries and regions around the world for which sponsors must follow a legal framework (see section 3). The countries/regions requiring the most extensive regulatory document public disclosure are the European Union (EU), Canada, Japan, and the United States (US). The EU and Canada both have requirements, Clinical Data Publication (CDP) and Public Release of Clinical Information (PRCI) respectively, that require Modules 2 and 5 (clinical reports) from marketing applications to be publicly disclosed [4, 5]. Japan's disclosure regulation requires a part of Module 1 and the entire Module 2 to be made public [6].

A brief comparison of the information available under these 3 policies:

Jurisdiction	Scope of the included submission documents
Japan	Part of Module 1 and the entire Module 2
EU	2.5, 2.7.1-4, Module 5 clinical reports with appendices of all studies
Canada	2.5, 2.7.1-4, Module 5 clinical reports with appendices of all studies

The US requires protocols and SAPs to be made public on Clinicaltrials.gov for all Phase 2 to 4 applicable clinical trials [7]. Additionally, the US has begun posting Assessment Aids for some trials that contain participant information [8].

For further information on worldwide national registries and databases, refer to the CTT Country Level Worksheet Supporting Documentation maintained by the PHUSE data transparency working group [9]. In addition to regulatory disclosures, many sponsors are voluntarily sharing data and documents for their clinical trials through formal research request processes. See Section 4, Existing Data Sharing Platforms.

3 Legal Framework

Regulatory documents might contain commercially confidential information (e.g., trade secrets and intellectual property) and personal information belonging to clinical trial participants and study personnel. Prior to disclosure, this information must be removed or transformed to protect the participants from being identifiable. This paper is primarily focused on the management of personal information. For more information on commercially sensitive information, please see the EFPIA Intellectual Property webpage [10].

Prior to publicly disclosing data and documents personal information of all stakeholders must be sufficiently anonymized (redacted and/or transformed) to prevent re-identification of participants and study personnel. Anonymization is defined as:

> The overall process of protecting the privacy of data subjects, including clinical study participants, and reducing the risk of re-identification by 1) modifying (e.g., suppressing, obscuring, aggregating, altering) identifiable information in structured data and documents, 2) assessing and controlling the residual risk of re-identification and 3) considering the context of the data release [11].

Modern clinical trials are often global, containing information from participant and study personnel from multiple parts of the world. A single clinical study report (CSR) is typically written to describe the global results of the trial and submitted to regulators in multiple regions. This single CSR may be publicly disclosed in each of the included regions and others despite conflicting privacy regulations and standards. Many sponsors attempt to anonymize and disclose their documents per the single most conservative privacy regulation being considered. This harmonized approach is not always successful, creating scenarios in which a document is anonymized differently for different regions. This could potentially result in increased trial participant re-identification risk.

3.1 Privacy Regulation

Often regarded as the most conservative privacy standard, the EU General Data Protection Regulation (GDPR) mandates that the principles of data protection do not apply to anonymized data. It defines anonymized data as "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable" (GDPR Page 5, Recital 26) [12]. It also states that, "to determine whether a natural person is identifiable, account should be taken of all the means reasonably used. . .either by the controller or by another person to identify the natural person directly or indirectly" (Page 5, Recital 26) [12]. The GDPR holds the data controllers of personal information responsible and liable for maintaining GDPR standards. Sponsors, as data controllers, must consider not just the possibility of re-identification utilizing the information within the documents being disclosed, but requires data controllers to consider the accessible information from other sources (e.g., voter registration records, healthcare data, social media) that could be cross-referenced with the publicly disclosed clinical trial documents in order to re-identify participants.

3.2 Data Disclosure Risk

The GDPR standard for anonymous data is often interpreted as an absolute standard – data either is or is not anonymous. Once data is anonymous, the informed consent of the participant is no longer required

for disclosure because the data cannot, in theory, be tied back to the participant. However, if the data are to be useful, experts agree there is still residual re-identification risk inherent in anonymized data [13]. Anonymization, in practice, is a process to ensure that the risk of re-identification is sufficiently low for its intended purpose – either public disclosure or disclosure directly to a researcher -- but the risk cannot be zero if the data are to retain utility.

Given the amount of information that is in the public domain, re-identification risk is increased with access to information about an individual or individuals in the dataset. With just birthdate, gender and a 5-digit zip code, research suggests that up 87% of the U.S. population is identifiable [14]. Some examples of personal information that is publicly accessible include:

- Voter registration records (can include name, address, race, gender, and birthday)
- On-line background check services (can include demographic and other personal information)
- Healthcare information bought and sold through health brokerage services
- Self-disclosed information on social media, television, and news publications

Most recently, COVID-19 vaccine news coverage presented sponsors with challenges for assessing re-identification risk. Participants volunteering for COVID-19-related products publicly shared demographic, location, and medical information (more specifically their medical history and medical experience associated with trial participation). Sponsors include statements about how sponsors share clinical trial data for voluntary and regulatory purposes in their participant informed consent forms. Participants often share their own information because they are proud of their role in advancing science and because they may want to influence others to participate in clinical trials. Sponsors acknowledge the right of the participants to disclose clinical trial participation details. However, these public details can be cross-referenced with clinical trial information that is publicly disclosed [15]. The ability to cross-reference what is publicly disclosed increases the trial participant's risk of being re-associated with data they may not want publicly disclosed. Regardless of what participants choose to disclose in the public domain, sponsors are legally bound as data controllers to ensure data and documents are sufficiently anonymized to prevent participant re-identification in regulatory public disclosure as well as voluntary disclosure to researchers.

4 Existing Data Sharing Platforms

4.1 Regulatory Sharing

Over the past decade, as global calls for transparency have increased, several regulations have been implemented that drive increases in data sharing through regulatory policy (see Section 2.2 Data Sharing Standards). As a result of these regulations, it has become easier for researchers and the general public to review and use the information contained in the full clinical study reports used in support of Marketing Applications of new products to Canada, the EU and Japan.
The publication of these documents is performed by the regulatory agencies via repositories that they own and manage. The repositories are open access to all users with limited restrictions. Per privacy experts, public disclosure requires an assumption that there is a 100% risk that re-identification will be attempted [16, 17]. This results in documents that are heavily redacted and transformed in order to

ensure clinical trial participants have a sufficiently low risk of being re-identified. The more significant level of redaction and transformation required for public disclosure in turn lowers the data utility of the clinical information, but the sponsors must make this trade-off in order to appropriately protect participant privacy in the context of public disclosure. The full datasets containing individual participant-level data are not published under any of these policies.
Starting in 2022 in the EU, there will be an additional database supporting transparency of clinical research data, the Clinical Trial Information System (CTIS) implemented in support of the EU Clinical Trial Regulation [18]. This system will make public additional information shared between regulators, countries, and sponsors during clinical study delivery. The full impact of the new regulation on transparency is still to be realized as the transitional option phase begins in early 2022. However, it should be noted that this system will only include information used to support interventional clinical research of drugs conducted in the European Union.

4.2 Voluntary Sharing

Beyond the legislation, clinical study sponsors across the globe have invested heavily in creating an ecosystem of tools, processes, and procedures to support clinical research and the advancement of medicines. This started with the Principles for Responsible Clinical Trial Data Sharing that were developed by EFPIA and PhRMA and signed by many across the industry in 2013. In Principle 1, "Biopharmaceutical companies commit to sharing upon request from qualified scientific and medical researchers' patient-level clinical trial data, study-level clinical trial data, and protocols from clinical trials in patients for medicines and indications approved in the United States (US) and European Union (EU) as necessary for conducting legitimate research." Since establishing this commitment, sponsors and non-sponsor organizations have invested time and resources towards creating data sharing platforms that include request processes, statistical analysis platforms, and scientific review boards to independently review requests. Additionally, sponsors have created systems and supported data transfers in support of legitimate research. This ecosystem has supported good science and is considered best practice in successful collaboration between industry and researchers to facilitate scientific and medical research and advancement across the globe [19].

The EFPIA Clinical Trial Data Portal Gateway contains a list of member companies' online portals with the aim of advancing responsible clinical trial data sharing [20].

Because voluntary data sharing has the following characteristics 1) takes place in secure environments with traceability, 2) the requestors are vetted to ensure they are using the data for a valid research purpose and 3) there are contractual controls in place (i.e. contractual requirements that the researcher does not attempt to re-identify individuals within the dataset and to limit the time period of access to the data), sponsors can assume a lower trial participant re-identification risk than the risk associated with regulatory public disclosure discussed in Section 4.1 Regulatory Sharing. Given a lower re-identification risk due to additional controls in place (ie. Data Sharing Agreements, controlled environments), sponsors can minimize redactions and transformation of data and documents provided under voluntary data sharing initiatives, resulting in greater data utility of the information provided through voluntary sharing.

4.2.1 Data Sharing Platforms

This section addressed two of the major data sharing platforms that are used by the global biopharmaceuticals industry. One of the benefits of these data sharing portals is creation of a central place for researchers to search for studies available for clinical data sharing by numerous sponsors. This should support easier access to available data. It should be noted that there are other data sharing platforms that have been established by government agencies, non-profit organizations, or other such collaborations. This paper cannot highlight all data sharing platforms that have been established due to the high change and growth in this area on an ongoing basis and will focus on two of such platforms. At a high level, data sharing platforms perform several key functions in support of both sponsors and researchers to enable data sharing:

- Receiving requests from researchers for data to be shared
- Providing workflow support to enable sponsors to manage the review of requests and record the decisions made during the review
- Providing an Independent Review Board to make decisions on the scientific value and soundness of proposals
- Providing an environment with statistical tools necessary to perform analysis of anonymized data shared for approved requests
- Supporting the execution of necessary data use agreements and ensuring all parties meet the obligations under the agreement (such as researcher publication)
- Providing statistics on the requests managed

Sponsors have data sharing policies that direct the approach to data sharing. Policies are typically aligned to baseline requirements established in the Principles for Responsible Clinical Data Sharing. These requirements enable sponsors to ensure appropriate diligence and stewardship of the clinical data assets that have been collected during global clinical research. It should be noted that sharing typically requires the following:

- Clinical trials in patients for medicines and indications that have been approved in the United States and the European Union.
- The requested purpose of research is considered to be legitimate to facilitate scientific and medical research.
- Requestors are qualified researchers and research teams include experienced biostatisticians.
- Requestors provide a statistical analysis plan that is considered of high quality by the scientific review boards.
- The ability to anonymize the data to ensure it can be shared appropriately according to global privacy laws and as committed to patients during informed consent (see details in Section 3, Legal Framework).
-

4.2.1.1 Clinical Study Data Request Portal

The Clinical Study Data Request Portal, also called CSDR, is an industry consortium to manage requests from researchers to biopharmaceutical companies. Clinical data approved for data sharing via the CSDR is often shared via the SAS Clinical Trial Data Transparency (CTDT) system, which is hosted by the SAS

Institute. A significant amount of data sharing has occurred through this portal, supporting progress in research. As of July 2021, over 3000 studies had been made available for sharing through this platform and through use of data on this platform there have been 84 separate publications [21].

4.2.1.2 Vivli

The Vivli organization is another data sharing platform and consortium (as of August 2021). Vivli is an all-in-one solution that provides a workflow request tool, support by an independent review board, and a technical environment to support the statistical analysis of the researchers. As of July 2021, Vivli includes over 6000 clinical studies that can be requested by qualified researchers and the research on approved data sharing via Vivli has resulted in 37 publications [22].

5 Summary

The biopharmaceutical industry is very committed to responsible data sharing while "safeguarding patient privacy, respecting the integrity of national regulatory systems, and maintaining incentives for investment in biomedical research" [19]. Since the launch of the principles in 2013, there have been many strides in harmonizing the way industry collects clinical trial data and reports trial results while satisfying diverse country regulations. Recent advances in technology have enabled pharmaceutical companies to implement the latest data anonymization processes. The pharmaceutical industry is widely recognised as leading the way in clinical trial data sharing with a continuing commitment to timely access to clinical trial results [23, 24).
Researchers are encouraged to request datasets directly from the sponsor's data sharing portals to realize the advantages of greater utility that comes with controlled and contractual data sharing policies.

References

1. International Council for Harmonisation, ICH Standards\CTD, M4: The Common Technical Document, accessed at: https://www.ich.org/page/ctd (last accessed 03 February 2022)
2. CDISC: CDISC Standards in the Clinical Research Process, accessed at: https://www.cdisc.org/standards (last accessed 03 February 2022)
3. International Council for Harmonisation, ICH Guidelines\All Guidelines, accessed at: https://www.ich.org/page/ich-guidelines (last accessed 03 February 2022)
4. EMA Clinical Data, Online access to clinical data for medicinal products for human use, accessed at https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication (last accessed 03 February 2022)
5. Health Canada, Health Canada finalizes regulations to provide public access to clinical information on drugs and medical devices accessed at: Health Canada finalizes regulations to provide public access to clinical information on drugs and medical devices - Canada.ca (last accessed 03 February 2022)

6. Hiroko Ebina, Jocelyn Colquhoun, Clinical trial disclosure landscape and awareness in Japan, accessed at https://journal.emwa.org/careers-in-medical-writing/clinical-trial-disclosure-landscape-and-awareness-in-japan/ (last accessed 03 February 2022)
7. US National Library of Medicine, ClinicalTrials.gov, accessed at: https://clinicaltrials.gov/ (last accessed 03 February 2022)
8. U.S. Food and Drug Administration, Assessment Aid, accessed at: https://www.fda.gov/about-fda/oncology-center-excellence/assessment-aid (last accessed 03 February 2022)
9. PHUSE: Clinical Trial Transparency Country-level Worksheet, accessed at https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/CTT+Country+Level+Worksheet+Supporting+Documentation.xlsx (last accessed 03 February 2022)
10. EFPIA, Intellectual Property, accessed at: https://www.efpia.eu/about-medicines/development-of-medicines/intellectual-property/ (last accessed 03 February 2022)
11. PHUSE: Terminology Harmonisation in Data Sharing and Disclosure Terms and Definitions, accessed at: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/WP063.pdf (last accessed 03 February 2022)
12. The European Parliament and of the Council, Regulation 2016/ 679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/ 46/ EC (General Data Protection Regulation), accessed at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 (last accessed 03 February 2022)
13. Mark Elliot, Elaine Mackey & Kieron O'Hara, The Anonymisation Decision-Making Framework 2nd Edition: European Practitioners' Guide, accessed at: https://msrbcel.files.wordpress.com/2020/11/adf-2nd-edition-1.pdf, (last accessed 03 February 2022)
14. Latanya Sweeney, Simple Demographics Often Identify People Uniquely, accessed at: https://dataprivacylab.org/projects/identifiability/paper1.pdf (last accessed 03 February 2022)
15. Kyle J McKibbin, Bradley A Malin, Ellen Wright Clayton, Protecting research data of publicly revealing participants, accessed at: https://academic.oup.com/jlb/article/8/2/lsab028/6352992, (last accessed 03 February 2022)
16. Institute of Medicine. 2015. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press. https://doi.org/10.17226/18998.
17. PHUSE: Calculating the Risk of Re-Identification of Patient-Level Data Using Quantitative Approach, accessed at: https://www.lexjansen.com/phuse/2016/dh/DH09.pdf (last accessed 03 February 2022)
18. European Medicines Agency, Clinical Trials Regulation, accessed at: https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trials-regulation (last accessed 03 February 2022)
19. PhRMA/EFPIA, Principles for Responsible Clinical Trial Data Sharing, accessed at: https://www.efpia.eu/media/25189/principles-for-responsible-clinical-trial-data-sharing.pdf (last accessed 03 February 2022)

20. EFPIA, Sharing clinical trial information, accessed at: https://www.efpia.eu/about-medicines/development-of-medicines/regulations-safety-supply/clinical-trials/sharing-clinical-trial-information/ (last accessed 03 February 2022)
21. ClinicalStudyDataRequest.com (CSDR) accessed at: https://www.clinicalstudydatarequest.com/Default.aspx (last accessed 03 February 2022)
22. Vivli, Center for Global Clinical Research Data, A global clinical research data sharing platform, accessed at: https://vivli.org/ (last accessed 03 February 2022)
23. EFPIA, EFPIA position on clinical trial transparency, accessed at: https://www.efpia.eu/news-events/the-efpia-view/statements-press-releases/efpia-position-on-clinical-trial-transparency/ (last accessed 03 February 2022)
24. Charles Piller, FDA and NIH let clinical trial sponsors keep results secret and break the law, accessed at: https://www.science.org/news/2020/01/fda-and-nih-let-clinical-trial-sponsors-keep-results-secret-and-break-law (last accessed 03 February 2022)

Appendix 1: Common Technical Document Clinical Modules and Associated ICH