PRA Health Sciences

WHITE PAPER

Data Sharing & Anonymization

Current Overview of Data Sharing within Clinical Trial Transparency

Authors

Cara Campora, PhD
Clinical Trial Transparency Manager
PRA Health Sciences

Laura Dodd, MSc
Manager, Global Clinical Trial Data Sharing
PRA Health Sciences

Executive Summary

Clinical trial transparency has emerged in recent years as a key function within pharmaceutical and academic organizations. Although trial transparency is typically associated with protocol registration and results disclosures, regulations and voluntary data sharing processes have led to a sharp increase in clinical trial data and documents now accessible to anyone with an internet connection. In particular, 42 CFR Part 11, which guides public disclosure of clinical information on ClinicalTrials.gov, as well as global regulatory changes to marketing authorization applications in Europe and Canada, such as European Medicines Agency (EMA) Policy 0070 and Health Canada Public Release of Clinical Information (PRCI), require sponsors to provide clinical documents that were previously not available to the general public. This evolution is a win for advocates of informed decision-making but poses operational and functional challenges in managing policy changes and different expectations between agencies. Sponsors preparing to share their clinical documents can incorporate best practices and implementation recommendations presented in this white paper, including the importance of planning and basic data anonymization strategies for protecting patient privacy while balancing data utility.

Key Data Sharing Drivers

Voluntary Initiatives

Effective trial transparency managers remain current on the latest information across different data sharing drivers, from voluntary commitments to legal requirements. With a focus on increased transparency, many pharmaceutical sponsors have established public disclosure policies that go beyond regulations, including manuscript publications, poster presentations at conferences, and access to data through specific agreements or portals. Several voluntary data sharing programs have emerged in recent years, such as the multi-sponsor request sites Vivli and ClinicalStudyDataRequest.com. Both platforms facilitate responsible sharing of patient-level data through a proposal system linking requesters to sponsor documents through a data sharing agreement process that includes an independent review panel to ensure confidentiality. The Duke Clinical Research Institute (DCRI) and collaborator Supporting Open Access for Researchers (SOAR) are other data sharing avenues, with common goals of increasing transparency and open research sharing to benefit future research tracks and improve patient access to information. Data repositories for specific therapeutic areas are also becoming more common, such as Project Data Sphere which provides access to oncology data. The International Committee of Journal Editors (ICMJE) has a data sharing policy stating that any manuscript submitted to ICMJE journals reporting results from clinical trials started after January 2019 must contain a data sharing plan in the protocol registration on a public registry. This incentivizes organizations and researchers to start considering internal data sharing strategies well before a clinical trial begins if they intend to report trial results in publications.

While many sponsors maintain their own platforms and policies for externally posting data, the voluntary aspect leads to significant variation and subjectivity in what is actually shared, with whom, and for what purpose the sponsor grants specific requests. These forward-thinking approaches improve consumer confidence, but require a shift in corporate planning, resourcing, and processes for accommodating ad hoc requests.

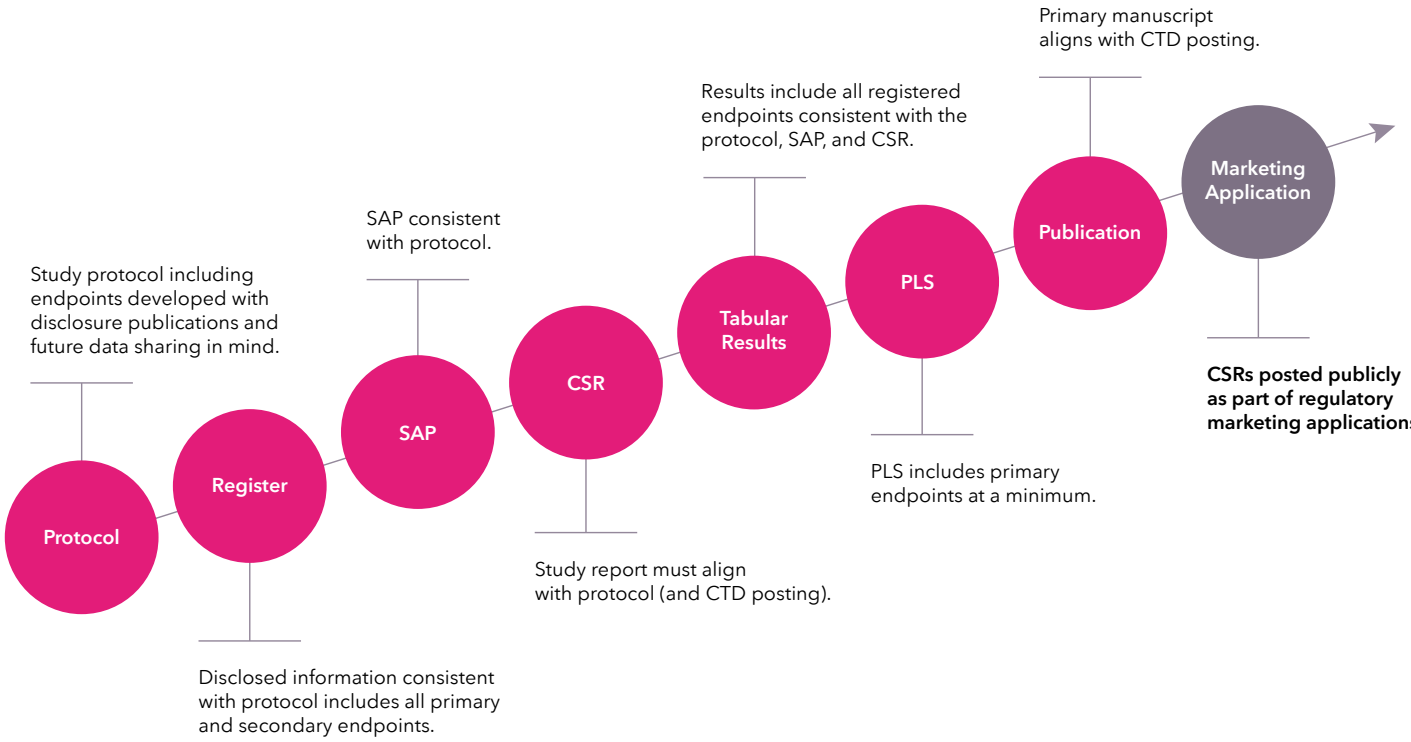
Sponsors best positioned with data transparency oversight incorporate voluntary data sharing into their regulatory obligations to complement their corporate planning and policies.

Regulations

Regulatory agencies worldwide have added to the transparency field in recent years by building on pre-existing guidance around public disclosure and enacting new policies to ensure that organizations are providing accurate, timely results following completion of clinical trials and as part of marketing applications. In particular, agencies in the United States, European Union (EU), Japan, and Canada promote responsible data sharing through regulations requiring the public release of protocols, statistical analysis plans, clinical study reports (CSRs), and related documents for improving transparency in the regulatory decision-making process and promoting better informed use of medicines. Within the disclosure lifecycle, data sharing activities generally occur as products become more mature, although future public disclosures should be taken into consideration in the early stages of planning and protocol development (Figure 1).

Food and Drug Administration and US 42 CFR Part 11

The Food and Drug Administration Amendments Act (FDAAA) 801 Final Rule, 42 CFR Part 11, effective in early 2017, provides clarification and new requirements that govern disclosure of clinical trial information on the ClinicalTrials.gov website, which is managed by the National Library of Medicine at the National Institutes of Health (NIH)[1]. 42 CFR Part 11 requires submission and publication of the latest amended protocol and statistical analysis plan as part of the results disclosure for applicable clinical trials with a Primary Completion Date on or after January 18, 2017[2]. Federally funded trials also require submission of a copy of the final individual consent form for eventual publication in a repository. NIH allows redaction of information that the sponsor deems necessary to safeguard personal protected data (PPD) and company confidential information (CCI) within these documents, which is a good opportunity for sponsors to establish a consistent approach to



CTD: clinical trial disclosures; CSR: clinical study report; PLS: plain language summary; SAP: statistical analysis plan.

Figure 1. Transparency Lifecycle and the Endpoint Continuum.

managing information intended for external audiences. The FDA also initiated a Clinical Data Summary Pilot Program in 2018 for assessing the feasibility of providing summary data from CSRs, which was concluded in March 2020 after obtaining feedback that will be used to refine potential future approaches that best harmonize disclosure of CSR and summary data between multiple regions and agencies[3].

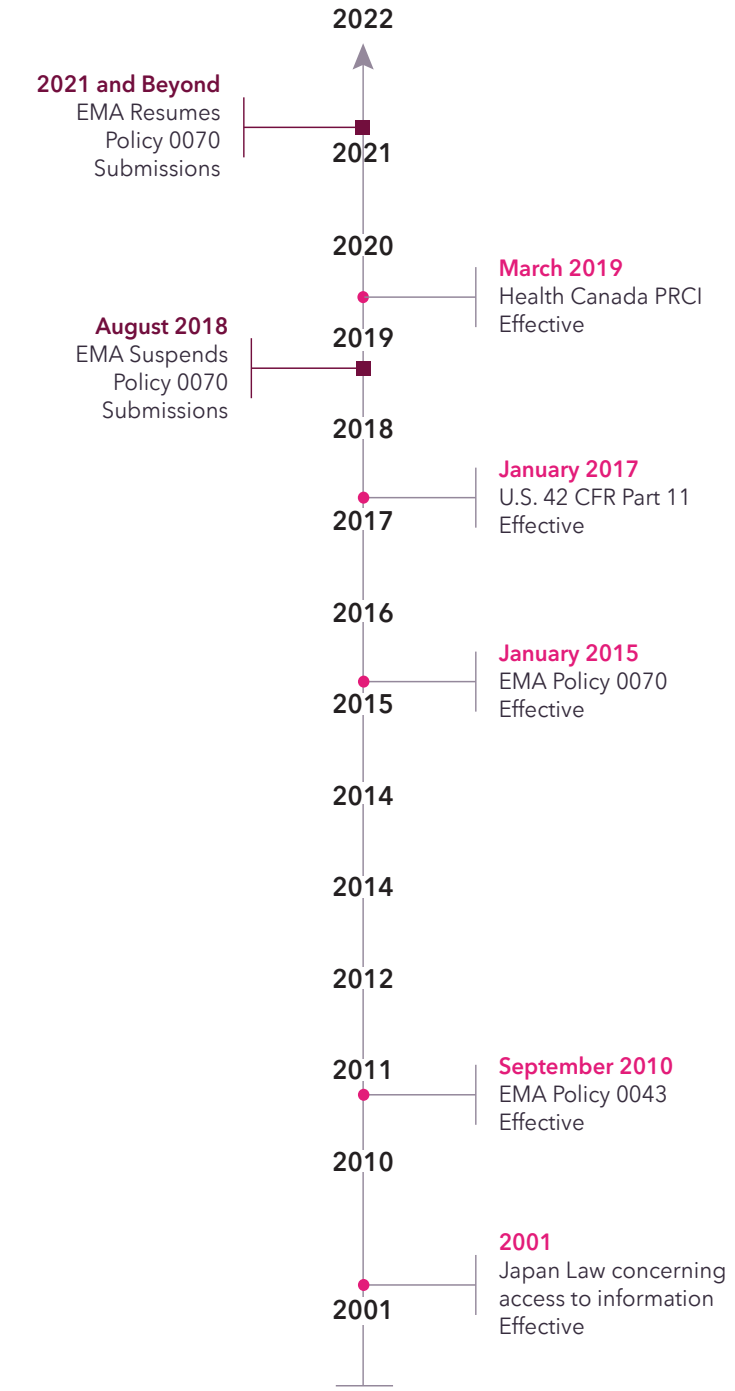
Japan Pharmaceuticals and Medical Devices Agency (PMDA)

Based on the Law Concerning Access to Information Held by Administrative Organizations in 2001, the public has the right to request certain types of documents that are submitted to national government organizations for products that have been approved in Japan[4]. While CSRs (Module 5) are not made public following approval of a Japanese New Drug Application, Module 2 (including overviews and summaries), part of Module 1 in the Common Technical Document (CTD) format submitted in Japanese as well as PMDA review reports are posted publicly on the PMDA's website. PMDA requests that sponsors prepare redacted versions of the CTD documents and review reports and to provide justification for PPD and CCI redactions, which can be rejected by the agency if the rationale for redaction is not acceptable[5].

European Medicines Agency

The European Medicines Agency (EMA) Policy 0070 launched in 2015 to publish clinical data and documents submitted by pharmaceutical sponsors in support of requests for marketing authorization in the EU[6]. Phase 1 of the policy includes clinical documents, while Phase 2 will eventually expand to include individual patient data, although the exact scope and timeline for implementing the second phase has not yet been detailed. The scope of Phase 1 includes clinical overviews and summaries, CSRs for individual clinical trials, and various supporting appendices, including the protocol, sample case report form, and statistical methods for all trials that support the marketing application. Given the sensitive nature of the data contained in these documents, EMA has encouraged sponsors to anonymize or redact PPD within clinical reports prior to submission to prevent re-identification of patients and study administrators. EMA allows CCI redactions on a very limited basis, as described in Annex 3 of the policy. Each proposed CCI redaction must be accompanied by a valid justification in table format and is subject to rejection by the EMA. Out of 1.3 million pages submitted during the first year of implementation, only 134 pages with

Timeline showing when Japan PMDA, FDAAA, EMA, and Health Canada regulatory requirements became effective.



CCI were accepted, or 0.01% of total pages published[7]. After a submission is accepted, all clinical documents that are deemed in scope as well as an anonymization report detailing the methodologies used to reduce the risk of participant re-identification are published on a publicly accessible clinical data portal. Over 80,000 document downloads were recorded for non-commercial purposes in the first year after studies were posted per Policy 0070, showing the increasing interest in access to clinical data and documents[7].

Although new submissions related to clinical data publication were suspended as of August 1, 2018, due to relocation of EMA headquarters and establishment of a new data portal, activities may resume in 2021, pending updated information from the EMA[8]. Retrospective information requests can be made via EMA Policy 0043, which allows interested parties to request redacted copies of certain documents that have been submitted following European Commission decision on marketing status of a product.

A separate EU Regulation 536/2014 details requirements for protocol registration and trial results[9]. This directive includes a new initiative to post plain language summaries with results, but does not require sponsors to post a protocol/SAP or CSR at the time of results posting until marketing authorization or up to seven years after the trial ends, whichever comes first[10].

	EMA Policy 0070	Health Canada PRCI	42 CFR Part 11	Japan PMDA
In Scope				
Agency Reviews	No	No	No	Yes
Summaries	Module 2	Module 2	No	Module 1 and 2
CSR	Module 5	Module 5	No	No
CRF	Yes	Yes	No	No
Protocol	Yes	Yes	Yes	No
SAP	Yes	Yes	Yes	No
IPD	Planned	Not Planned	No	No
De-Identification Method	Redaction or Quantitative Anonymization	Quantitative Anonymization Preferred	Redaction	Redaction
Terminology & Redaction Color Coding	PPD, light blue with black overlay and CCI, black with red overlay	PI and CBI; color scheme not specified	No overlay text; color not specified (black as default)	No overlay text; color not specified (black as default)
Timeline	Proposed package due between day 181 and 220*	Proposed package due within 60 days of decision	Due at the time of results disclosure on ClinicalTrials.gov	Posted after NDA approval
Supplemental Documents	Justification tables, anonymization report	Redaction control sheet, anonymization report	Tabular results	Justification report for redactions
Consultations	One meeting per sponsor; evaluates all redactions and methods; can reject CCI	Process Initiation Meeting at start of process; evaluates all redactions and methods; can reject CCI	None	None; evaluates all redactions and methods; can reject CCI

CBI: company business information; CCI: company confidential information; CRF: case report form; IPD: individual patient data; NDA: new drug application; PI: personal information; PPD: protected personal data; SAP: statistical analysis plan

* If there is an EMA backlog, the deadline for packages is determined based on a notification letter from EMA and can be six months to one year or longer[13].

Table 1. Key Similarities and Differences between Regulatory Data Sharing Drivers

For more information regarding clinical trial registrations and results disclosure, read PRA's whitepaper here and plain language summaries white paper here.

Health Canada

Health Canada's Public Release of Clinical Information (PRCI) policy became effective in March 2019, with a total of 41 dossiers published within the first year[11,12]. This initiative provides public access to clinical information that allows independent analysis of data and supports new scientific research directions. The Health Canada guidance is similar to EMA Policy 0070 in terms of overall scope and processes, although there are several key differences. PRCI prefers that patient data in study reports is anonymized based on a quantitative risk assessment rather than using qualitative redaction to maintain data utility, submission timelines are significantly shorter, and terminology is slightly different (Table 1). Despite these differences, Health Canada allows submission of documents processed for Policy 0070 that have already been accepted by the EMA. While the guidance is geared toward prospective applications, retrospective requests can be made through the clinical information portal for documents that were submitted prior to the 2019 effective date.

Protecting Participant Privacy

A critical aspect of data sharing involves protecting participant information. As the push to publicly release more clinical data has come to the forefront, so have global concerns about data privacy. Ensuring that information release is governed by best practices to reduce the risk of re-identifying individuals represented in various data sets is a key focus for data controllers. In what may seem to be a conflicting end result, sponsors and regulators are tasked with finding the delicate balance between protecting participant privacy through redaction or anonymization strategies and providing data utility, which can be defined as the degree to which a reader can interpret and make meaningful interpretations about the information.

PPD in clinical documents can include direct and indirect identifiers for both study participants and study staff. Direct identifiers are unique to a specific individual, such as full name, phone number, email address, or social security number. Indirect identifiers are parameters that alone might not pose a high risk of re-identification, but can be used in combination with other information to potentially identify a specific individual, such as city, state, zip code, event dates, demographic parameters, and sensitive medical information.

Example comparing data utility. Redaction leads to a significant loss of information, while quantitative anonymization transforms the data to retain generalized information based on a risk assessment.

Redaction

Patient ID	Age	Country	Event Date
9800-130	24	USA	14-Oct-14
9800-225	65	Japan	12-Dec-15
9801-301	15	USA	22-Aug-14
9801-556	92	Belgium	1-Jul-14
9802-446	33	Italy	3-Mar-15

Patient ID	Age	Country	Event Date

Quantitative Anonymization

Patient ID	Age	Country	Event Date
9800-130	24	USA	14-Oct-14
9800-225	65	Japan	12-Dec-15
9801-301	15	USA	22-Aug-14
9801-556	92	Belgium	1-Jul-14
9802-446	33	Italy	3-Mar-15

Patient ID	Age	Country	Event Date	Cycle Day
4455-786	20-30	North America	24-Oct-14	67
4455-987	55-65	Asia	22-Dec-15	492
3457-101	<18	North America	11-Sep-14	23
9801-556	>89	Europe	1-Nov-14	74
3458-224	30-40	Europe	23-Mar-15	217

Redaction is a qualitative method of applying a colored box over direct and indirect identifiers related to an individual. This type of physical masking can be manually applied or semi-automated and is attractive at first glance due to perceived cost effectiveness and availability of common software tools. It is most suitable for short documents that do not contain a lot of PPD. In documents that contain a significant amount of PPD, redaction offers little to no data utility in that all information is fully masked and decisions about what elements to redact or retain can be subjective. This approach often leads to visually fragmented documents that are difficult to read and draw conclusions upon.

As an alternative to the limitations of redactions, anonymization strategies based on quantitative risk assessments are trending as an evolving best practice for data sharing. Anonymization transforms data variables from an original raw state to generalized or banded categories so that the reader understands the context while statistically reducing the risk of individual re-identification. For example, participant event dates like date of death or medical event can be offset or replaced by cycle day, and exact ages can be re-grouped to age bands showing general information. Using the popular data anonymization approach called "k-anonymity," data sets and related data within documents can be transformed based on the maximum probability of 1/k of being re-identified. Generally, maximum risk thresholds of 0.09 (k=11, 1/k), or a 9% probability of an individual in the data set being re-identified, are acceptable by regulatory agencies based on prior research and precedents in the field of data privacy and public release[14]. Validated software tools can efficiently calculate risk assessments, establish transformation strategies, and implement anonymization techniques to transform data simultaneously across multiple clinical datasets and documents, saving time on manual efforts and ensuring confidence in the output by using an automated, quantitative approach supported by robust quality control steps.

Planning for Success

With ever-changing policies, new initiatives, and increased demand for public access to documents, the magnitude of managing both voluntary and legal requirements can be challenging for even the most established organizations. Advanced planning and setting up multifunctional teams to support data sharing tasks are essential in this dynamic field.

Sponsors should consider assigning a staff member within their transparency and disclosures or regulatory affairs department as the primary point person to track incoming data sharing requests and prospective marketing applications. The point person can be responsible for tracking projects, coordinating meetings with vendors, and liaising between team members to facilitate review processes according to internal standard operating procedures.

In addition to a trial transparency manager, sponsor review team members typically consist of a representative from each of the following departments:

- Intellectual property or patent lawyer to review for CCI
- Clinical team member to provide input on risk assessments, PPD, and CCI
- Regulatory affairs to ensure compliance with timelines and content
- Biostatistics to provide input on risk assessments
- Other roles, such as a data privacy officer, depending on internal processes.

Establishing a small, core team of well-informed personnel is key to efficiently responding to data sharing initiatives. Internal teams can assist in creating a standardized approach or set of rules used to manage the different PPD identifiers and CCI elements within all documents to maintain consistency across various disclosure activities. It is important that all review team members are aware of the strict justifications required to pass proposed CCI redactions through both EMA and Health Canada reviews and have realistic expectations about the level of disclosure required by regulators in support of marketing applications. Although disclosure activities are not often considered at the start of a new project, proactive planning by medical writing and clinical trial teams with future public

Successful data sharing submissions require advanced planning with team leads experienced with navigating global requirements and facilitating a standardized approach

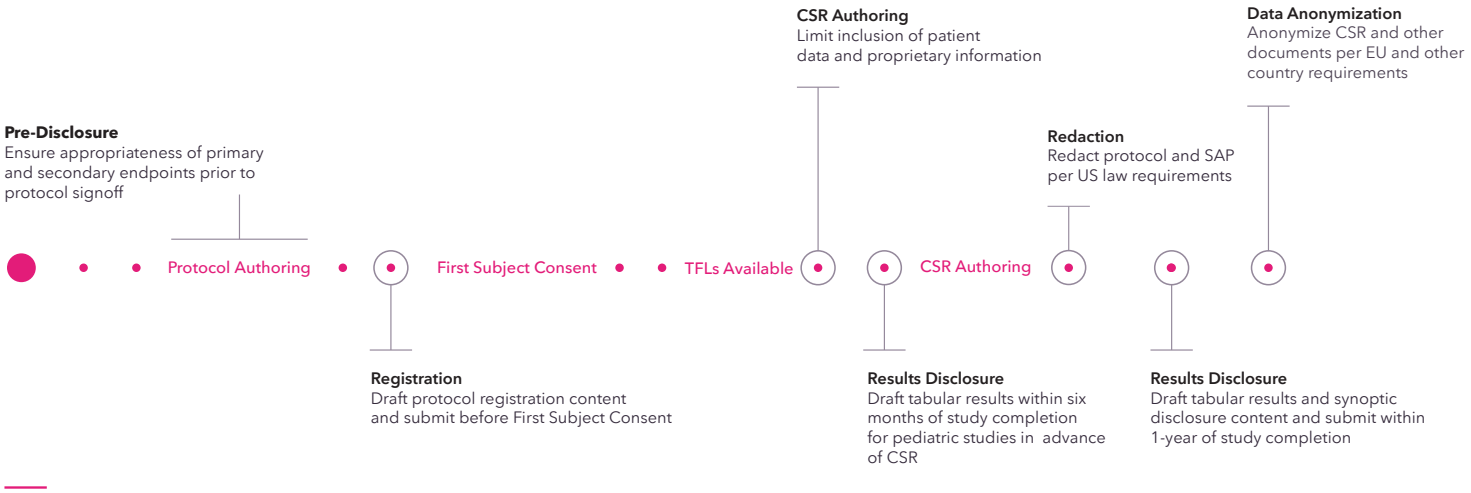


Figure 2. Transparency Lifecycle and Clinical Trial Timelines

disclosures in mind can significantly reduce the downstream workload and cost by excluding unnecessary PPD and CCI within clinical documents (Figure 2). All review team members within a sponsor organization should keep in mind alignment with other public sharing activities so that documents in the public domain are consistent to the extent possible.

Given the many nuances involved in disclosure of clinical documents, many organizations prefer to work closely with a well-established external company with experience in this niche area. Much of the heavy lifting around preparing the documents for public release, such as redaction and anonymization, can be outsourced to a reliable vendor that has all of the validated software tools needed to accurately anonymize data and documents to reduce the risk of re-identification and guide the team through the complexities of the different policies.

Conclusion

Building successful, comprehensive data sharing strategies within organizations takes planning and patience. Staying current on data sharing initiatives, establishing informed internal teams, and having tools in place to implement best practices in anonymizing participant data are important steps to seamlessly managing retrospective and proactive requests to disclose clinical data and documents. PRA utilizes anonymization software with experienced, cross-functional staff for successful execution of even the most daunting data sharing tasks while maintaining the highest standards of safeguarding participant privacy and commercially confidential information.

References

1. Food and Drug Administration Amendments Act (FDAAA 801) Final Rule (42 CFR Part 11). https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission
2. Protocol Registration and Results System Final Rule Information website. https://prsinfo.clinicaltrials.gov
3. FDA Clinical Data Summary Program website. https://www.fda.gov/drugs/developmentapprovalprocess/ucm589210.htm
4. Pharmaceutical Administration and Regulations in Japan, 2019. http://www.jpma.or.jp/english/parj/pdf/2019.pdf
5. Ebina, H. and Colquhoun, J. 2019. Clinical trial disclosure landscape and awareness in Japan. Medical Writing, Volume 28 Number 1.
6. European Medicines Agency Policy on Publication of Clinical Data for Medicinal Products for Human Use. https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf
7. European Medicines Agency Clinical Data Publication (Policy 0070) Report Oct 2016-Oct 2017. https://www.ema.europa.eu/en/documents/report/clinical-data-publication-policy-0070-report-oct-2016-oct-2017_en.pdf
8. Billiones, Raquel. 2020: Will the EU clinical trials regulation be implemented this year? https://www.linkedin.com/pulse/2020-eu-clinical-trials-regulation-implemented-year-raquel-billiones/
9. EU Clinical Trials Regulation 536/2014. https://ec.europa.eu/health/human-use/clinical-trials/regulation_en
10. Appendix on disclosure rules to the "Functional specifications for the EU portal and EU database to be audited – EMA/42176/2014." 2015. https://www.ema.europa.eu/en/documents/other/appendix-disclosure-rules-functional-specifications-eu-portal-eu-database-be-audited_en.pdf
11. Guidance Document on Public Release of Clinical Information website. https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release-clinical-information-guidance.html
12. McGuinness, N. 2020. One Year In: How the Opening of Health Canada's Portal Affects You. https://privacy-analytics.com/de-id-university/one-year-in-how-the-opening-of-health-canadas-portal-affects-you/
13. Eibert, S.M. 2018. Policies 0070 and 0043: Juggling different requirements. Medical Writing, Volume 27 Number 2.
14. El Emam, K. 2013. Guide to the De-Identification of Personal Health Information. CRC Press (Auerbach).

Contact Information

If you would like more information or to discuss PRA's Data Sharing and Anonymization services, please contact us at GRATrialTransparency@prahs.com, or:

Erik Lakes, MSc, MSc.RA
Senior Director, Clinical Trial Transparency
LakesErik@prahs.com

Cara Campora, PhD
Clinical Trial Transparency Manager
CamporaCara@prahs.com

World Headquarters
4130 ParkLake Avenue, Suite 400
Raleigh, North Carolina 27612 USA

Phone: +1 (919) 786 8200
Fax: +1 (919) 786 8201
www.prahs.com