CRDSA

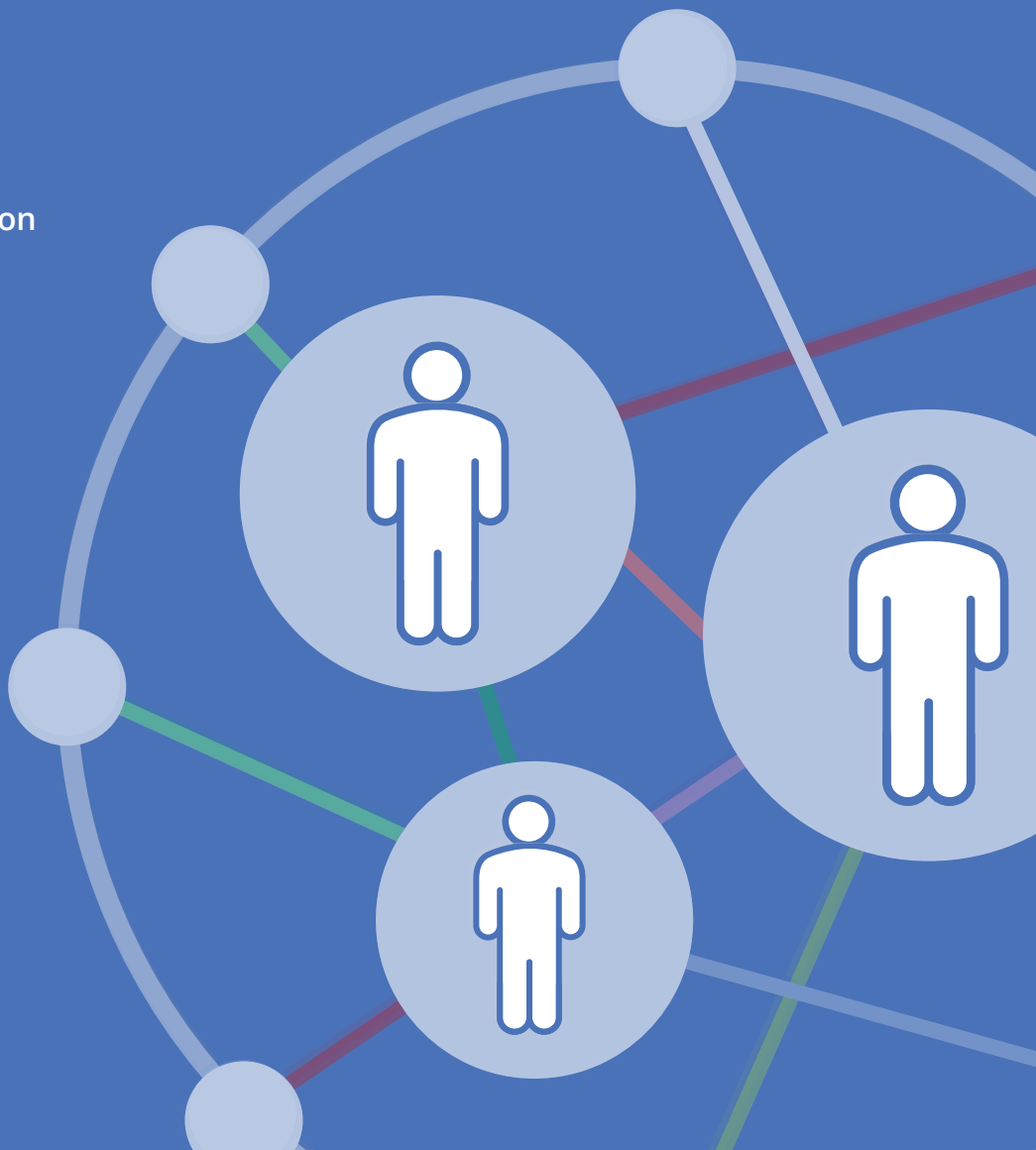
Clinical Research Data Sharing Alliance

A Review of Biopharma Sponsor Data Sharing Policies and Protection Methodologies (2023)

Version: 2.0

Work Group: Data Protection

Date: 28 September 2023



A Review of Biopharma Sponsor Data Sharing Policies and Protection Methodologies (2023)

Liz Roberts (UCB), Luk Arbuckle (Privacy Analytics), Andrei Belcin (Privacy Analytics), Cathal Gallagher (D-Wise), Janine Hodgson (Takeda), Aaron Mann (Clinical Research Data Sharing Alliance) on behalf of the CRDSA Data Protection Work Group

Check for Updates »

Executive Summary

This whitepaper updates CRDSA's 2022 systematic review of data contributor policies and data protection methodologies. This update includes information published by 35 biopharma sponsors across three data sharing platforms: Vivli, Clinical Study Data Request (CSDR), and the Yale Open Data Access (YODA) Project. We focused on policy elements impacting end-user research utility, including what data is shared and how data will be transformed to protect patient privacy.

There was a year-on-year net increase of 6 biopharma data contributors included in the 2023 analysis, an increase of 20.7% over 2022. This growth in data sharing through multi-contributor platforms is a significant benefit for researchers, making it easier for them to request and use data from multiple contributors. For data contributors, multi-sponsor platforms deliver a managed data sharing and use process that is well understood by the research community.

Small data contributors (under 5,000 employees) and midsize contributors (5,000 to 25,000 employees) showed a substantial year-on-year improvement in their commitment to share across the dataset and supporting documentation categories in the analysis. In 2023, commitments now exceed 80% for all types of datasets and supporting documentation (with the exception of small-contributor sharing of dataset specifications at 71%). This is an average year-on-year increase of 19% for midsize contributors and 36% for small contributors. This consistency of dataset- and documentation-sharing policy, across sponsors of all sizes, establishes a reference benchmark for all data contributing organizations.

This report also highlights where work is needed, specifically regarding how the data is being transformed in the contribution process. The areas of data protection methodology and data transformation transparency continue to pose challenges.

Given the progress shown in this report, we are confident that these challenges will be addressed, resulting in a data sharing ecosystem that more efficiently connects researchers with highly usable data and also benefits data contributors, because they can be more confident that the shared data will actually be used. This is a win-win-win scenario — a win for researchers, a win for data contributors, and most importantly, a win for patients who rely on researchers to identify new scientific insights that ultimately could lead to new treatments.

Introduction

This is an update to CRDSA's whitepaper published in September 2022. This revision includes results from a systematic review, using the same methodology, that was conducted in August 2023.

Sharing patient-level data generated from clinical trials is fundamental to the advancement of science and the improvement of public health. The use cases for shared data are well established and include novel clinical trial design and enrichment strategies, predictive preclinical and clinical models, clinical trial simulation tools, biomarkers, clinical outcomes assessments, and more.[1] Shared clinical research data also has the power to transform the trial process itself, improving the patient experience and delivering life-saving and life-changing therapies faster and at less cost to society.

However, when sponsors put data sharing into practice, there can be high variability in both contribution volume and utility to end users. It is important to recognize the challenges faced by data contributors. Ensuring that contributions maximize research utility must be balanced with the equally important need to responsibly protect patient privacy. This balancing act results in a wide spectrum of contribution approaches, some of which may compromise data utility to the point where scientifically interpretable analyses become increasingly challenging.

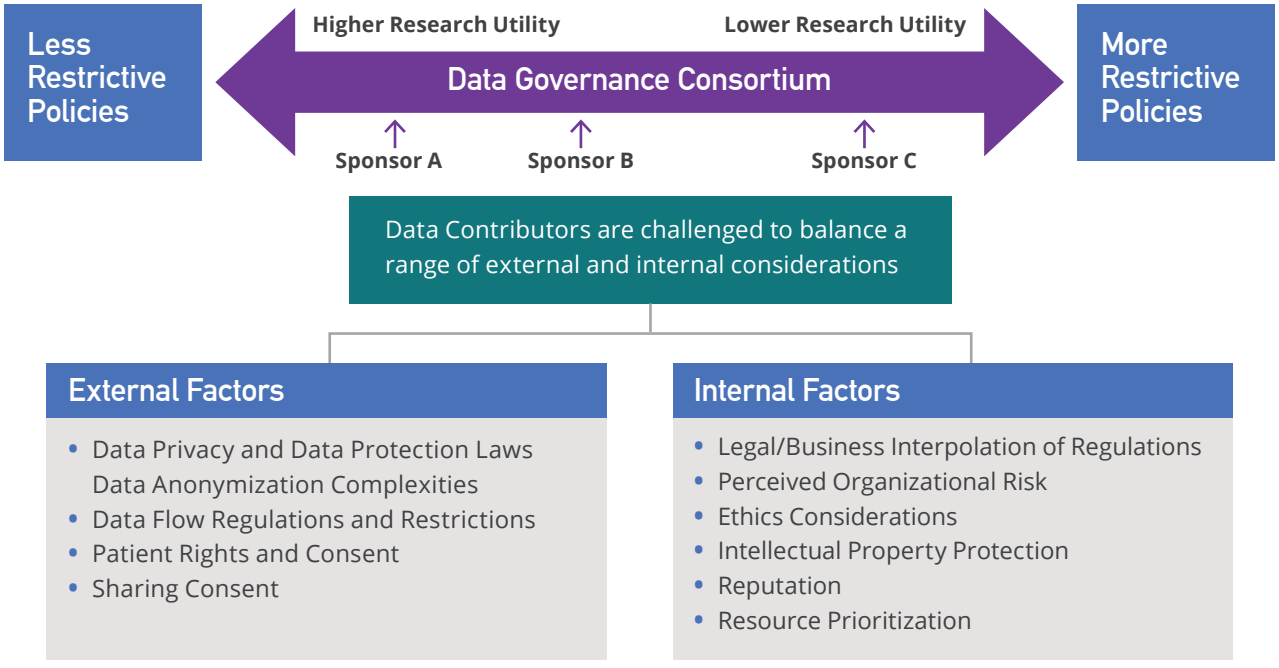


Figure 1: Data Governance Continuum

The results of a survey of academic and biopharma researchers published in CRDSA's "Establishing a basis for secondary use standards in clinical trials"[2] indicated that the datasets and supporting documentation contributed are critical determinants of research utility. They provide researchers with the context and underlying information needed to adequately understand the datasets being used for analysis. The survey also indicated the importance of transparency in the data protection methodology being applied to transform the data for secondary use.

This whitepaper focuses on elements impacting end-user research utility, including what data is shared and how data will be transformed to protect patient privacy. In conducting the initial 2022 systematic review, the authors aimed to create meaningful benchmarks to help guide data contributor policy development. This update reports on policy elements that have improved over the course of the year and highlights areas where continued work is needed.

Methodology

Source Data

The source data for this review was collected from publicly available information published by trial sponsors across three data sharing platforms: Vivli, Clinical Study Data Request (CSDR), and the Yale Open Data Access (YODA) Project [Appendix A]. The information that sponsors publish on these platforms can include informational statements, data points (e.g., lists of supplied documents), and linked policy and/or process documents.

Of the sponsors listed on these platforms, in 2023, 35 biopharma sponsors provided information sufficient for this analysis, while most academic sponsors did not provide detailed information. To provide meaningful comparisons, this review focuses on the information provided by the biopharma sponsors.

The authors are aware that some of the sponsor information may be out of date. However, it was determined that our analysis would rely on publicly available information published on the data sharing platforms as of our review periods (April/May 2022 and August/September 2023). Critically, the published sponsor information is what is available to researchers. We encourage sponsors to review available information to ensure it is consistent with their current policies and practices.

Data Protection Assessment Methodology

In addition to synthesizing the information supplied, the team developed a consistent methodology to categorize data protection approaches. The objective of applying this methodology was to determine whether a sponsor uses a risk-based data protection approach (i.e., objectively supported through measures of identifiability, or the risk of re-identification, and a reasonableness standard). A risk-based approach to data protection is generally thought to improve the end-user utility of data contributions[3].

In our analysis, we encountered divergences in what is considered risk-based anonymization, similarly shown by Rodriguez et al,[4] who conducted a literature review on recommended clinical trial anonymization (or de-identification) approaches to determine what was consistent across the recommendations made. Because of these divergences, there is currently no single standardized set of recommendations related to anonymizing clinical trial datasets for sharing. Industry groups, including TransCelerate[5] and PHUSE,[6] have proposed frameworks, while ISO/IEC 27559 [7] offers a new industry-agnostic standard.

The role of controlled access on the degree of anonymization, especially the effect on risk-based anonymization, is depicted below by the SAFE Data Standard,[8] published after the literature review with the intent of promoting standardization and efficiency in the sharing of clinical trial data:

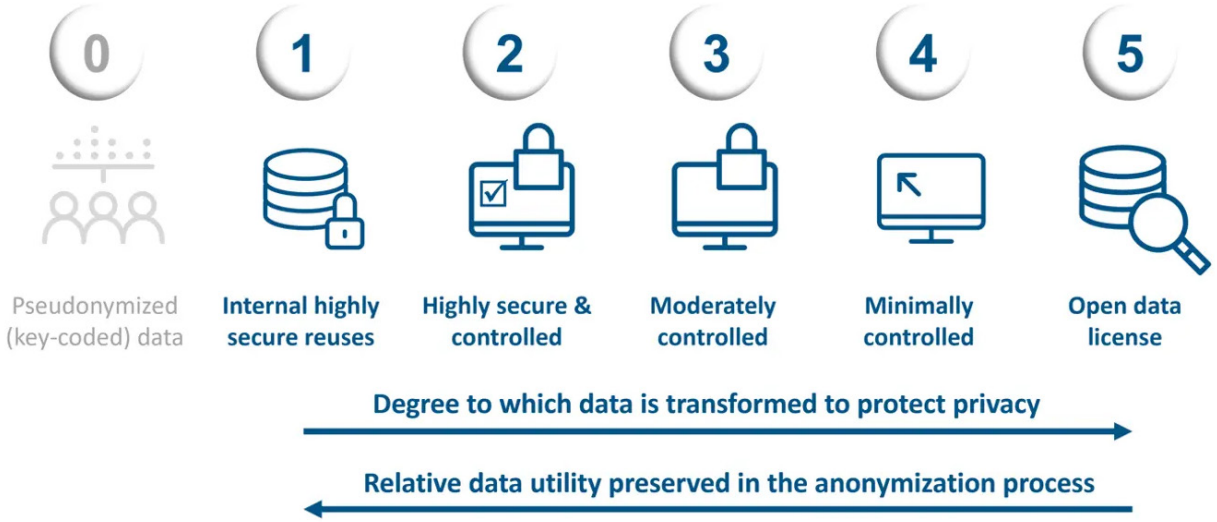


Figure 2: *SAFE data rating by Bamford et al. (Applied Clinical Trials, 2022):* Description of the SAFE Data Standard rating system, where data rated from 1 to 5 has been anonymized to reflect the context of data disclosure, with an increasing degree of data transformation and associated impact on data utility. "Internal" refers to the trial sponsor reusing its data.

The data rating applied informs the degree of anonymization required to ensure a re-identification probability at or below a threshold. Platforms, including those in this review (Vivli, the YODA Project, and CSDR), may rate at 2 on the scale above (depending on implementation). Internal use within the sponsor organization is often rated at 1, whereas case-by-case partnerships with individual organizations (not sharing platforms) often score a 3. To classify approaches with confidence, the authors scaled expectations of the sponsors' published standards to reflect a rating of 2 (i.e., highly secure and controlled) for the platforms included in this review.

From the publicly available anonymization standards by pharma sponsors on Vivli, CSDR, and the YODA Project sharing platforms, we can observe three criteria for evaluating a standard to classify it as "risk-based" (in decreasing order of confidence):

1. The extent to which an approach to anonymization is modified based on specific factors influencing risk: e.g., study population, disease prevalence, data sensitivity, system controls, context risk, various re-identification attack scenarios, and adversary profiles
2. The explicit claim by the sponsor that their anonymization approach is "risk-based"
3. A mention of a "risk assessment" in addition to a specified rule set

These criteria were determined from both domain knowledge and the tendencies in specifications among a given sponsor's published standards scaled to the expected access controls of the sharing platforms. The review used the following categorizations in assessing the data protection methodology used by a sponsor:

- **Criterion 1** is classified as using a risk-based anonymization approach, and we can also infer the use of a quantitative approach.
- **Criterion 2** is classified as using a risk-based anonymization approach.
- **Criterion 3** is possibly a risk-based anonymization approach.
- **Else,** if a sponsor uses a rule set minus the risk assessment component, then we infer use of a rules-based approach.

Collation and Interpretation

There is inherent variability in the underlying information used in this review. As noted above, some information may not be current. It was also noted that sponsors might use terms differently and/or interchangeably (e.g., anonymization and de-identification). Further, our analysis, particularly regarding the applied data protection methodology, is based on a best-efforts interpretation of the supplied information. For these reasons, the authors have chosen to present this review at the aggregate level, tiered by sponsor size.

The direct and indirect (interpreted) information was collated into three tiers based on sponsor size. Tiers were determined using the total sponsor employee count as follows:

Tier 1	Tier 2	Tier 3
25,000 employees and above	**5,000 to 24,999 employees**	**4,999 employees or fewer**

The breakpoints between tiers were distinct, and no sponsors were within 13% of an adjacent tier. The list of reviewed sponsors is provided in Appendix A.

Results

There was a year-on-year net increase of 6 biopharma data contributors included in the 2023 analysis, a 20.7% increase over 2022.

Table 1: Number of Sponsors by Tier

	Tier 1 25,000 and above		Tier 2 5,000 to 24,999		Tier 3 4,999 or fewer	
	2022	2023	2022	2023	2022	2023
Number of Sponsors	12	14	11	14	6	7

A notable improvement seen in the 2023 analysis is the consistency across tiers. In 2022, larger tier 1 sponsors were substantially more likely to share datasets and documentation than tier 2 and tier 3 sponsors. In 2023, commitments now exceed 80% across all tiers for all types of datasets and supporting documentation, with the exception of data specifications in tiers 1 and 3.

Table 2: Sharing of Datasets and Documentation

	Tier 1 25,000 and above		Tier 2 5,000 to 24,999		Tier 3 4,999 or fewer	
	2022	2023	2022	2023	2022	2023
Raw	100%	100%	73%	86%	83%	100%
Analysis	92%	93%	82%	86%	67%	86%
Protocol	100%	100%	73%	93%	83%	100%
Annotated CRF	100%	93%	64%	86%	67%	86%
Reporting and Analysis Plan / SAP	100%	100%	73%	86%	67%	86%
CSR	92%	86%	82%	86%	33%	86%
Data Specifications	75%	79%	64%	86%	50%	71%
Average	**94% v. 93% (nominal decrease)**		**73% v. 87% (19% increase)**		**64% v. 88% (36% increase)**	

(See Appendix B for additional resources, including information on dataset and documentation types.)

It is worth noting that, while the stated commitment to share clinical study reports exceeds 80%, the authors are aware that some sponsors are sharing the synopsis rather than the complete CSR. This variability applies across all tiers, including large sponsors. It is important that sponsors make this distinction as part of their stated policy, and we encourage all sponsors to share the complete CSR with researchers.

Data Protection Methodology

As our research centered around public domain documentation from sponsors, our ability to assess the details was limited to what was explicitly available. The authors did not bring forward any personal or informal knowledge of current practices by sponsors. However, based on confidential knowledge among the group of authors, we believe that many sponsors have outdated documentation describing their anonymization standards. Exacerbating the situation, there are still a number of sponsors not providing any documentation on their anonymization approach.

Table 3: Data Protection Methodology

	Tier 1 25,000 and above		Tier 2 5,000 to 24,999		Tier 3 4,999 or fewer	
	2022	2023	2022	2023	2022	2023
Term: anonymized	83%	79%	92%	100%	50%	71%
Term: de-identified	8.5%	21%	0%	0%	33%	29%
Uses both interchangeably or missing field	8.5%	0%	8%	0%	0%	0%
Risk based	42%	36%	27%	21%	17%	14%
Rules-based sharing following a similar data protection methodology*	42%	43%	45%	50%	33%	43%
HIPAA/other approach	17%	21%	27%	29%	50%	43%
Details provided	67%	64%	55%	57%	50%	43%

* Where additional detailed documentation was provided, Appendix C summarizes the general methodology used solely as the rules-based approach or in conjunction with a risk assessment.

Few sponsors stated that their assessment of re-identification (disclosure) risk involved various factors influencing risk (Criterion 1), of which we infer to include some form of quantitative measurement of risk and in turn informs the amount of transformation on the trial data to achieve an expected utility by the data recipients (e.g., researchers, regulators).

Some sponsors provided a short document (2 to 3 pages) that makes claims about their approach being risk based (Criterion 2) without sharing the set of transformations considered and the basis for choosing to enact them on the clinical trial data at a high level.

Similarly, many sponsors generally use the phrase "risk assessment" (Criterion 3) preceding the outline of their rules-based anonymization, without any link to what the assessment entails and how it affects the degree to which rules are applied to trial data. Regardless, we gave sponsors the benefit of the doubt that they may, in fact, be using a risk-based approach.

Otherwise, all other approaches are classified as rules based, in which there is no evidence, no claim, and no risk assessment to even suggest that a risk-based approach is used.

Additional Dimensions

Sponsors also provide information about their general data sharing policies regarding what studies will be shared, when studies become eligible for secondary use data sharing, and exceptions. As above, the authors collated the information by sponsor-size tier. While large sponsors tend to provide somewhat more context to the information presented, we did not find significant policy differences between tiers. Therefore, these results are presented as applicable across tiers.

What studies are shared?

Most sponsors share Phase 2–4 interventional studies (although some provide no information). Eleven sponsors also state that they share Phase 1 studies; however, it should be noted that Phase 1 studies may fall under a study exception (see next page) because of the limited number of patients. Generally, sponsors limit sharing to studies that have been through a regulatory approval process (typically in one or more of US, EU, and Japan) or if the development product is terminated.

When are studies shared?

Of the 35 sponsors reviewed, 30 provide information on timelines for study sharing eligibility. This is a 13% increase in information provided compared to 2022. However, only 18 sponsors specify a numeric time frame, while the rest provide generic statements (for example, "after approval" with no specified timeline) or do not provide any sharing timeline information.

Table 4: Study Share Timelines

Sponsor Statement	n=35
12 months after results, approval, and publication	11%
18 months after approval or termination and study report completion and publication	23%
Within 6 months of publishing and approval in US and EU	17%
Generic statement (e.g., after approval and after publication)	34%
No information provided	15%

Across the 35 sponsors, there were, broadly speaking, 5 different approaches to communicating sharing timelines. There were no meaningful distribution differences between the tiers. In the absence of a consensus approach, the authors encourage sponsors to adopt Good Pharma Scorecard's criterion of making data available by 6 months after approval by the FDA or EMA or 18 months after a trial's completion date, whichever is later.[9]

Study Exceptions

Study exceptions were generally consistent across sponsors. The most common reason study data cannot be shared is an inability to achieve an anonymization threshold that adequately protects patient privacy. This can be because of a small patient population (typically under 50), rare disease indications, or geographic considerations (e.g., a single-center trial). Studies that have been co-developed or are co-owned are also frequently cited as out of scope.

Data Exceptions

Data exceptions were also generally consistent across sponsors, including imaging (x-rays, MRI scans), genetic data, exploratory biomarkers, and non-English documents. It should be noted that generally accepted best-practice data protection methodology for some data types, including images and genetic data, is still being developed.

Discussion and Recommendations

With the 2022 whitepaper, the authors aimed to create meaningful benchmarks to help guide data contributor policy development. This 2023 update allows us to look at areas where the policy landscape has evolved and, equally important, where more work is needed.

The 20% year-on-year increase in the number of biopharma sponsors is important to recognize. This growth in sponsor data sharing through multi-contributor platforms like Vivli, the YODA Project, and Clinical Study Data Request (CSDR) is a significant benefit for researchers, making it easier for them to request and use data from multiple contributors. For data contributors, multi-sponsor platforms deliver a managed data sharing and use process that is well understood by the research community. Additionally, using the SAFE data rating framework, these platforms are considered "highly secure and controlled," helping to ensure the responsible processing and use of data contributions.

As stated in the introduction, establishing policy benchmarks was a goal of the 2022 review. In 2023, the stated policies on Datasets and Documentation are substantially more consistent across the tiers, creating a reference benchmark for sponsors of all sizes.

The scope of sharing was reasonably consistent across data contributors. Datasets and associated documents from Phase 2–4 interventional clinical trials (sometimes also Phase 1 trials) were generally in scope for sharing. However, the study sharing timelines and associated requirements (e.g., publication, approval, termination) were much less consistent across sponsors, with nearly half providing no information or supplying general statements with no timeline commitments.

The types of protections applied to the data prior to sharing and how the risk of re-identification was mitigated were much less clear. This is critical information to share with researchers so they can determine whether the data will be available in a form that allows analyses to proceed as planned—i.e., for the shared data to have sufficient clinical utility.

Recommendations

1. Clarity of Available Information

- The terms "de-identified" and "anonymized" are sometimes used synonymously, although they have different meanings in different regions. Alignment of terminology and definitions would provide much-needed clarity across the global data sharing ecosystem.
- Data contributors should provide more information up front explaining their approach to data protection. This is especially important when researchers plan to combine data from multiple contributors and need clarity to understand whether such data can be pooled. For example:
    - » Will any data, such as rare adverse events, need to be removed prior to sharing? If this information is known, then researchers whose analysis requires such data would be able to identify such trials as potentially not being able to provide sufficient data utility.

- » If the research requires granular demographic data, is it clear how these variables will be transformed (or whether even present) in the datasets that are to be shared?
- » If the research is in an area where seasonality is an important consideration, how will this have been addressed in the datasets that are to be shared?

- Some contributors describe the extent to which their approach to data protection is modified based on specific factors, including study population, disease prevalence, data sensitivity, system controls, context risk, various re-identification attack scenarios, and adversary profiles. This best-practice approach provides much more clarity to researchers and is also aligned with the preferred approach by regulators as part of mandatory document publication policies.

Sponsors: Recommended Anonymization Methodology Detail

Overall, the authors recommend that sponsors across tiers provide the following details as baseline components in their anonymization standards on clinical trial data sharing platforms:

- Specificity on the risk assessment, application of quantitative or qualitative methodology, and the relevant factors considered in the assessment
- The way in which the risk assessment informs the transformations of the clinical trial data to render it anonymized
- References to, or in support of, the methodology used to anonymize and produce useful data for end users

As a result, the recipients of the data can both form their own judgment on how confident they are in the data meeting patient privacy requirements and build an expectation of the utility of the data. As standards emerge and validation of data integrity becomes more common, end users will be more likely to trust the data and insights derived, improving the coordination of health research and increasing public trust.

2. Benchmarking

- The responsible sharing of patient-level data is not intended to be a box-ticking exercise conveying compliance with various regulations and guidelines. To support meaningful data sharing, data contributors must ensure that processes are followed to protect individuals' data privacy but also retain as much data utility as possible in the resultant datasets and documents that are to be shared. If there is insufficient transparency (data are not "FAIR"[10]), then the potential secondary benefit is lost. If requested data sources do not contain sufficient data utility, then time and money are lost by both the research team and the data contributor. It is also possible that the omission of that data could have deleterious effects on meta-analyses, including drawing incorrect inferences based on incomplete or nonrepresentative data sets.

- Benchmarking is an important tool for data contributors, as it allows organizations to understand where they stand in the data sharing ecosystem. It can provide information related to return on investment, including:
  - How many (and which) clinical trials are requested (and how often)
  - Whether the data in these trials can be used to generate new research as planned (e.g., Does it contain the required utility? Do research projects using contributed data tend to be completed, or are they abandoned?)
  - Knowledge of the new scientific insights that could lead to new potential treatments

Conclusion

Our updated analysis shows substantial progress in data contributor policy commitments around the sharing of datasets and supporting documentation. Addressing what is being shared is an important step towards a data sharing ecosystem that efficiently connects researchers with highly usable data and provides the information and context needed to ensure responsible research use.

This analysis also highlights where work is needed, specifically regarding how the data is being transformed in the contribution process. This requires addressing challenges including the privacy methodology being applied, dataset-level processing transparency, and the process harmonization needed to ensure a FAIR data environment.

By working together, the research and data contributor communities (recognizing that many organizations are active in both fields) can create a paradigm shift within the data sharing ecosystem that more efficiently connects researchers with highly usable data and also benefits data contributors, as they can be more confident that the shared data will actually be used. This is a win-win-win scenario — a win for researchers, a win for data contributors, and most importantly, a win for patients who rely on researchers to identify new scientific insights that ultimately could lead to new treatments.

About CRDSA

CRDSA is a multi-stakeholder alliance that serves the clinical data sharing ecosystem. Our mission is to accelerate the discovery and delivery of life-saving and life-changing therapies to patients by expanding the research value of secondary use data. Broad access to these data has the power to transform the research process, improve trial design and delivery, and benefit the patients who donate their time and their data as part of the clinical development process. To find out more please visit crdsalliance.org.

References

1. Karpen SR, White JK, Mullin AP et al. Effective data sharing as a conduit for advancing medical product development. *Ther Innov Regul Sci.* 2021;55(3):591–600. doi:10.1007/s43441-020-00255-8

2. Odame E, Burgess T, Arbuckle L et al. *Appl Clin Trials.* Establishing a basis for secondary use standards for clinical trials. March 8, 2023. https://www.appliedclinicaltrialsonline.com/view/establishing-a-basis-for-secondary-use-standards-for-clinical-trials

3. Polonetsky J, Tene O, Finch K. Shades of gray: seeing the full spectrum of practical data de-identification. *Santa Clara L Rev.* 2016;56:593–629. https://digitalcommons.law.scu.edu/lawreview/vol56/iss3/3

4. Rodriguez A, Tuck C, Dozier MF, et al. Current recommendations/practices for anonymising data from clinical trials in order to make it available for sharing: A scoping review. *Clin Trials.* 2022 Aug;19(4) doi:10.1177/17407745221087469. Epub 2022 Jun 22.

5. TransCelerate BioPharma Inc. Privacy methodology for data sharing. n.d. Accessed September 26, 2023. https://www.transceleratebiopharmainc.com/initiatives/privacy-methodology-for-data-sharing-2/

6. PHUSE Advance Hub. Data transparency. August 17, 2023. https://advance.phuse.global/display/WEL/Data+Transparency

7. International Organization for Standardization. ISO/IEC 27559:2022(en): Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework.2022; Accessed September 26, 2023. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:27559:ed-1:v1:en

8. Bamford S, Lyons S, Arbuckle L, Chetelat P. Sharing anonymized and functionally effective (SAFE) data standard for safely sharing rich clinical trial data. *Appl Clin Trials.* 2022;31(7/8). https://www.appliedclinicaltrialsonline.com/view/sharing-anonymized-and-functionally-effective-safe-data-standard-for-safely-sharing-rich-clinical-trial-data

9. Axson SA, Mello MM, Lincow D, et al. Clinical trial transparency and data sharing among biopharmaceutical companies and the role of company size, location and product type: a cross-sectional descriptive analysis. *BMJ Open.* 2021;11:e053248. doi:10.1136/bmjopen-2021-053248

10. Wilkinson M, Dumontier M, Aalbersberg I et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016(3)160018. doi:10.1038/sdata.2016.18

Appendix A

List of Platforms and Sponsors

Platforms

- Clinical Study Data Request (CSDR) (https://www.clinicalstudydatarequest.com/)
- Vivli (https://vivli.org/)
- The Yale Open Data Access (YODA) Project (https://yoda.yale.edu/)

Sponsors

- AbbVie (https://www.abbvie.com/)*
- Alnylam (https://www.alnylam.com/)
- Astellas (https://www.astellas.com/en/)
- AstraZeneca (https://www.astrazeneca.com/)
- Bayer (https://www.bayer.com/en/)
- Biogen (https://www.biogen.com/en_us/home.html)
- Boehringer Ingelheim (https://www.boehringer-ingelheim.com/)
- Bristol Myers Squibb (https://www.bms.com/)
- Chugai (https://www.chugai-pharm.co.jp/english/)*
- Daiichi-Sankyo (https://www.daiichisankyo.com/)
- Eisai (https://www.eisai.com/index.html)
- Galapagos (https://www.glpg.com/)*
- Ipsen (https://www.ipsen.com/)*
- Lilly (https://www.lilly.com/)
- GSK (https://www.gsk.com/en-gb/)
- Grunenthal (https://www.grunenthal.com/)
- Johnson & Johnson (https://www.jnj.com/)
- Kyowa Kirin (https://www.kyowakirin.com/index.html)
- Lundbeck (https://www.lundbeck.com/us)
- Mitsubishi Tanabe Pharma (https://www.mt-pharma.co.jp/e/)
- Novartis (https://www.novartis.com/)
- Ono (https://us.ono-pharma.com/)*
- Otsuka (https://www.otsuka-us.com/)
- Pfizer (https://www.pfizer.com/)
- Regeneron (https://www.regeneron.com/)
- Roche (https://www.roche.com/)
- Sanofi (https://www.sanofi.com/)
- Shionogi (https://www.shionogi.com/us/en/)
- SpecGx LLC, a subsidiary of Mallinckrodt Pharmaceuticals (https://www.mallinckrodt.com/)
- Sumitomo Pharma/Sunovion Pharmaceuticals, Inc. (https://www.sumitomo-pharma.com/)
- Taiho Pharmaceutical (https://www.taihooncology.com/)
- Takeda (https://www.takeda.com/)
- Tempus (https://www.tempus.com/)^
- Teva (https://www.tevapharm.com/)*
- UCB (https://www.ucb.com/)
- ViiV (https://viivhealthcare.com/)*

* new for 2023
^ 2022 only

Appendix B

Additional Resources and References

CRDSA

Webinar: Mind the Data Sharing Gap: Navigating sponsor policies and data protection methodologies https://www.youtube.com/watch?v=HdL5kMP_Ugs

Paper: Establishing a Basis for Secondary Use Standards in Clinical Trials https://www.appliedclinicaltrialsonline.com/view/establishing-a-basis-for-secondary-use-standards-for-clinical-trials

Webinar: Data Sharing Technologies: Introducing a User-Centric Framework https://www.youtube.com/watch?v=N_RuHjjg1PU

Vivli

Webinar: Applying the SAFE Data Standard to Securely Share Clinical Trial Data https://youtu.be/WS1PysSpj3s?si=wR9n6q64OiQSV8Yj

Video: Why is the Data Anonymized/De-Identified Prior to Sharing? https://youtu.be/PzX5cMCQ3XI

Video: What are the Supporting Documents Provided Along with the IPD? https://youtu.be/dBE1hUDvWb8

Video: What is a Clinical Study Report (CSR)? https://youtu.be/ozRJBMJOWBI

Yale Open Data Access (YODA) Project

Reference: Community Data Sharing Resources

https://yoda.yale.edu/about/community-data-sharing-resources/

Article: Sharing clinical trial data: lessons from the YODA Project https://www.statnews.com/2019/11/18/data-sharing-clinical-trials-lessons-yoda-project/

Appendix C

Table 3: Data Protection Methodology

Where sponsors provided additional detailed documentation, the following summarizes the general methodology used solely as the rules-based approach or in conjunction with a risk assessment.

1. Remove personally identifiable information from the dataset of the 18 identifiers (as defined by HIPAA US).
2. Recoding identifiers and research subjects' identification code numbers.
3. Removing free text verbatim terms.
4. Replacing date of birth by age (banded).
5. Replacing all original dates relating to a study subject using either: a) dummy date method or b) study day offset method.
6. Reviewing and removing/redacting other PII: sites/labs, locations, investigators, imaging data (MRI, x-ray), etc.
7. Quality control checks on the anonymization and packaging of the data/documents to be shared in separated locations from the original data.
8. Destroying link (key code) between de-id data and source trial data, storing anonymized data separately from the source data, and erasing remnants of processing the source data.