

Being a “good data citizen”: Are you contemplating sharing your clinical trial data?

Sowmya Gabbula, Principal Statistical Programmer, Clinical Solutions Group - an IQVIA business, Sowmya.Gabbula@iqvia.com
Krishna Chotneeru, MPH., Associate Director – Data Science & Statistics - Alnylam Pharmaceuticals, kchotneeru@alnylam.com

Background

Why Share – Being a “Good Data Citizen”

Aligning With Your Senior Leadership

Creating A Transparency Governing Body

Industry organizations and regulatory agencies are encouraging companies and research institutes to share patient level data sets:

- EU Policy 0070 Phase II will require release of data sets in the future; Health Canada has stated that they will implement a similar policy when best practices and mode of transfer are decided.
- EFPIA PhRMA and BIO Clinical Trial Transparency principles include sharing individual patient level data and data sets.
- ICMJE requires sponsors to include a data sharing “plan” when submitting articles for publication.
- Clincltrials.gov, following on ICMJE’s requirement, now contains a question asking whether data will be shared, which is publicly visible in every registry record (as of 2018)

Data sharing is an opportunity to engage with the research and patient communities. Voluntarily making *anonymized* patient-level data sets and supporting documentation for certain studies available to interested researchers for re-analysis.

Why share data?

- Demonstrates commitment to science and the research community
- Demonstrates commitment to patients
- Can help accelerate discovery
- Allows sponsors to be prepared for impending requirements
- Vast majority of patients feel there would be no consequence to sharing their data
- Sponsors support what goes out and to whom through a proposal vetting and data contract process

Meet specifically with your senior leadership with a proposed data sharing strategy to implement a data sharing program

- Provide justification as a Good Data Citizen
- Be prepared to discuss potential costs as well as speak to the unpredictability of the request volume – you may receive few requests or a tidal wave.
- Ensure you have support from other key stakeholders who will participate in the process – Legal, Clinical Research, Medical Affairs, Data Sciences, etc.

Form a working or governing body comprised of key representatives from Legal, Clinical Research, Data Sciences, Corporate Communications, Medical Affairs, Clinical Operations, and Compliance and any other important stakeholders. Use this group to:

- Develop a Clinical Trial Transparency and/or Data Sharing Policy
- Select a data sharing platform (vendor) or commit to creating one internally (resource heavy)
- Create and finalize criteria for defining what studies would be in scope for data sharing
- Initiate and oversee collaboration between Data Sciences and Transparency on a data de-identification pilot to de-identify at least one data package
- Develop a process for preparing data, identify a repository for de-identified data, and develop internal spot check/QC process
- Consider applying process to any internal needs/requests for de-identified data

Leveraging a Data Sharing Platform

Data Sharing Platforms are independent consortiums that facilitate the responsible sharing of patient-level data using global data sharing and analytics systems.

Their goal is to promote and coordinate trial data sharing and re-use, providing a space for researchers to request data and for sponsors to safely release data for re-analysis.







How does it work?

Preparing data for de-identification

How is data de-identified?

- Per Sponsor’s data sharing policy, parameters are defined for responsible data sharing requests and what data/documentation can be shared.
- Researchers must provide a proposal that is deemed of scientific merit by the sponsor.
 - ✓ Proposals are reviewed and approved by an internal review board
 - ✓ If approved, researcher signs a data use contract
 - ✓ Data is robustly anonymized by sponsor or vendor and uploaded to a secure, password-protected environment.
- Researcher analyses the data within the data environment
 - ✓ Data should not be downloadable/removable from the system
 - ✓ Analysis results only can be downloaded

Depending on the data sharing request and agreed upon criteria

- Select the data package of interest
- Provide RAW (SDTM) and / or Analysis (ADaM) data
- Provide full or partial data package, per requirements of data use agreement
- Provide all the pertinent documents such as, Protocol, Analysis Plan, Case report forms (annotated), Data specification and Data Definition files.

Direct Identifiers

Subject number, Unique Subject Identifier, Subject Initials, Sequence numbers, etc.

Quasi-Identifiers (indirect)

Sex, date of birth/age, Date of death, Geographic location, Ethnicity, Race, schooling, Marital Status, Event dates, etc.

Transformations:

- Direct identifiers undergo masking
- Quasi-identifiers can be generalized, date-shifted (per PhUSE guidance, for dates) or suppressed.

Variable	Final strategy
Age	10-year interval
Country	Masked
Region group 1 (5 groups)	Masked
Region group 2 (3 groups)	Masked
Sex	Retained
Race	Masked
Ethnicity	Masked
Date of birth	Masked
Date of death	Shifted
Height	Retained
Weight	Retained
AE start, end dates	Shifted
Hospitalization admission date	Shifted

De-identification QC Checks

Developing a Data Sharing request review framework

References

These checks will be done against the datasets which have been de-identified:

Reviewing the meta data such as

- Number of files received;
- Number of variables within domains
- Number of records

Do a check on the transformations based on the de-ID specification document such as:

- Masking & Suppressing- Direct Identifiers (FPE) Form preserving encryption
- PHUSE date shifting, global suppression, parameter level suppression, generalization - Quasi identifiers
- Global suppression of the data value: Non-direct or Quasi identifier such as QNAM in SDTM.SUPPxx

Objective: Establish a reviewer group for the purpose of reviewing and evaluating data request proposals

Members should include:

- Regulatory strategy
- Program/Product Strategy
- Clinical Research
- Legal
- Medical Affairs
- Safety
- Data Sciences

Leverage existing Review Board for data sharing request review if/where possible

Data sharing platforms:

- <https://www.clinicalstudydatarequest.com/Default.aspx>
- <https://yoda.yale.edu/>
- <https://vivli.org/>
- <https://archive.bio.org/articles/bio-principles-clinical-trial-data-sharing>
- <https://www.efpia.eu/media/25189/principles-for-responsible-clinical-trial-data-sharing.pdf>
- <http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>
- https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-1.pdf