An Analysis of Data De-Identification Practices in Use at Leading Biopharmaceutical Companies

Dave Handelsman, Senior Director of Industry Strategy
dave.handelsman@d-wise.com
Research Triangle Park, North Carolina, USA

Introduction

Clinical trial data transparency initiatives have rapidly evolved across the biopharmaceutical industry.

Critical to the successful implementation of these transparency solutions is the ability to share data that:

- Has been properly de-identified in order to protect patient confidentiality
- **AND**
- Complies with national regulations
- **AND**
- Does not diminish data utility or potential analysis value

This poster will provide a comparison between published company-specific de-identification rules, highlight the differences and similarities, and ultimately describe recommendations for industry.

Methodology

ClinicalStudyDataRequest.com identifies eleven companies that have taken a collaborative approach to clinical trial data transparency. Of these, eight have published their de-identification strategies.

These public strategies have been compared and the results have been summarized regarding the similarities and differences across this select group.

This de-identification analysis does not directly take into account ongoing industry initiatives that are in progress and not yet published.

Analysis – Similarities

Data Type	De-Identification Action
Personally identifiable information (PII)	Remove
Unique subject ID	Re-code
Subject ID	Re-code
Initials	Remove
Investigator names and locations	Remove
Age > 89	Remove age and categorize as Age >=90
Dictionary coded terms	Retain

Analysis – Differences

Data Type	De-Identification Approaches
Date of Birth	• Delete month and day while keeping the year • Remove the entire date value • Replace the value with a calculated age
Date fields, except for Date of Birth	• Calculate study day from designated anchor point for each patient • Apply a random offset, per patient, to each patient's date value
Free-text values	• Review manually and redact selectively • Always remove
Patients per Site, Sites per Study, Sites per Country	• Aggregate sites with few* patients • Aggregate countries with few* sites and/or patients • Avoid de-identifying single-site trials * "Few" patients is frequently defined as <10; "Few" sites are not defined
Variable names	• Review for location information • Ignore

Conclusions

Significant Variation in Industry

Clinical trial data transparency and sharing, and its companion process of patient-level de-identification, have been addressed with historic speed in the biopharmaceutical industry. Not surprisingly, comprehensive and consistent industry-driven guidelines, rules and best practices still need to be established.

Ongoing Collaborations

Formal industry collaborations regarding de-identification are in process. The PhUSE De-Identification Working Group has recently closed comments on its draft de-identification rules, and the final document is expected to be published in Spring 2015. This document is focused exclusively on the de-identification of CDISC SDTM 3.2, however, and will only serve as a rough guide for data that conforms to other CDISC models, corporate standards, or data lacking standards conformance. Separately, TransCelerate is expected to publish its model approach to de-identification and anonymization later in 2015.

Recommendations

As patient-level de-identification matures, a well-defined set of de-identification rules directly lends itself to a more streamlined approach than the time-consuming and resource-intensive SAS programming process in use today.

A robust software solution that can apply industry rules and provide flexibility to handle both historic and modern clinical trial data will be of tremendous benefit in addressing this immature problem space. Such a solution will need to be designed to evolve with the emerging business and regulatory needs of the industry, while never losing sight of the vital need to protect patient privacy.

References

clinicalstudydatarequest.com/Documents/Astellas-Data-Anonymization-Standards.pdf
clinicalstudydatarequest.com/Documents/Bayer-Data-Anonymization-Standards-Apr-14-2014.pdf
clinicalstudydatarequest.com/Documents/Anonymisation of Clinical Trial Datasets-BI.pdf
clinicalstudydatarequest.com/Documents/Anonymisation of Clinical Trial Datasets.pdf
clinicalstudydatarequest.com/Documents/Novartis-Global-Data-Anonymization-Standards.pdf
clinicalstudydatarequest.com/Documents/Sanofi-DeIdentification-Guide.pdf
clinicalstudydatarequest.com/Documents/Takeda-Anonymization-Standard-10Oct-2014-Final.pdf
clinicalstudydatarequest.com/Documents/UCB-Anonimization-Clinical-Trial-Datasets.pdf

d-Wise
Moving Beyond the Data