Striking the Right Balance Between Transparency and Privacy in Clinical Trial Data Sharing

Clinical trial researchers must balance the at times contradictory needs for transparency, with protecting participants' privacy and looking after companies' intellectual property. Transparency and information sharing can be voluntary and motivated by the desire to progress science through collaboration. However, increasingly sharing of trial documents and data is driven by regulatory requirements. The different global regulations ensure that the information released is governed by best practices which reduce the risk of re-identifying individuals represented in various data sets. In what may seem to be a conflicting end result, sponsors and regulators must find the delicate balance between protecting participant privacy, through redaction or anonymisation strategies, and providing data utility. This can be defined as the degree to which a reader can analyse and make meaningful interpretations from the information.

So how can pharmaceutical or academic researchers perform this delicate balancing act? And when operating in multiple geographic regions how can they meet the expectations of different regulatory bodies? In this article we outline the requirements of different regulators. We suggest ways that researchers can assess risk, plan and take steps that maximise data utility while meeting the expectations of regulators and patients.

Redaction vs Transformation

Redaction is a method of masking information by applying a box over direct or indirect identifiers. It can be done manually or semiautomatically using most common software tools. Because of this it can be perceived as attractive and cost-effective. For short documents with little personal information (PI) or Protected Personal Data (PPD) it may be the logical choice. However, it has little to no data utility since all of the information is fully hidden. In addition, deciding what to redact can be subjective.

Transformation is the process of pseudonymising, offsetting or generalising direct or indirect information relating to participants. Direct identifiers can be the full name, subject numbers, phone number, email address or a government ID number. Indirect identifiers alone might not lead to reidentification, but combined with other information could be used to identify an individual. They include city, state, demographics and sensitive medical information. In transformation, instead of including a participant's age, participants' ages would instead be banded into groups. Validated software tools can assess the risk of reidentification, establish transformation strategies and implement anonymisation techniques. This can save time by automatically applying anonymisation to multiple clinical datasets and documents. This automated quantitative approach, when combined with robust quality control steps, can ensure confidence in the outputs.

The Challenges of Voluntary Data Sharing

Many sponsors already share data beyond the regulatory requirements with several voluntary data sharing programmes established in recent years. The objectives are varied but include contributing to open research to benefit future research studies and improving patients' access to information. Because the drivers are voluntary, decisions about subjectivity and other factors are varied and can be inconsistent. What gets shared, with whom, and for what purpose may be open to interpretation by the different sponsors and stakeholders. From an organisational perspective, it calls for a change in corporate planning, resourcing and putting in place processes to deal with information requests.

Regulatory Requirements

Regional regulatory bodies also have similarities and differences and different timelines for reporting information. These can be as detailed as which colour to use for overlays of PPD and confidential business information (CBI) or company confidential information (CCI). The Food and Drug Administration (FDA) and National Institute of Health allow redaction of information that the sponsor deems necessary to safeguard PPD and CCI. In Japan, the Pharmaceuticals and Medical Devices Agency (PMDA) requests justification for PPD and CCI redactions. These can be rejected by PMDA if the rationale for redaction is not acceptable. The European Medicines Agency (EMA) allows CCI redaction on a very limited basis and each proposed redaction must be accompanied by a justification which may be rejected. During the first year of the implementation of Policy 0070, of 1.3 million pages submitted just 134 pages, or 0.01% of the total pages published were accepted. A separate EU regulation includes an initiative to post plain language summaries alongside the final results summaries at the time of marketing authorisation. Health Canada's Public Release of Clinical Information (PRCI) policy provides public access to clinical information that allows independent analysis of data and supports new scientific research directions. The guidance is similar to EMA's Policy 0070 in overall scope and processes.

Reidentification Risk

If data are insufficiently anonymised there is a significant risk of reidentification. In a court case between the Southern Illinoisan and The Department of Public Health, expert witness Dr Latanya Sweeney successfully reidentified 18 out of 20 individuals in a neuroblastoma data set from the Illinois cancer registry. For the remaining two individuals, the witness was able to suggest one of two alternative names. While a graduate student at MIT, Dr Sweeney, identified the Governor of Massachusetts's medical information using birth date, gender and ZIP code information and a publicly available database. In Canada, a national broadcaster aired a report of the death of a 26 year old student taking a particular drug. The student was reidentified using information from the adverse drugs reaction database released by Health Canada. Aside from the loss of privacy and distress to individuals or their families, these reidentifications can lead to lawsuits and pay outs.

Strategies to Minimise Risk and Maximise Research Success

Managing both voluntary and regulatory requirements is challenging for any research organisation. Advance planning, agreed standardised processes and clearly defined roles will help to ensure the correct measures are established and upheld. A small team of well-informed cross-functional members is important to respond to data sharing initiatives. Proactive planning for the required disclosure activities should be built in from start up. Regulators look for justification for anonymising both PPD and CCI, and planning report writing in advance can save time later on. In addition, many of the tasks related to anonymisation and document preparation for release can be outsourced to relieve pressure on clinical teams. Having validated software tools and knowledgeable people to implement them can help you to avoid the potential pitfalls and ensure that your findings will contribute to future research.

REFERENCES

1. SOUTHERN ILLINOISAN v. ILLINOIS DEPARTMENT OF PUBLIC HEALTH (2006) | FindLaw
2. L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570
3. Khaled El Emam, Fida Kamal Dankar, Protecting Privacy Using k-Anonymity, Journal of the American Medical Informatics Association, Volume 15, Issue 5, September 2008, Pages 627– 637, https://doi.org/10.1197/jamia.M2716

Sarah Johnston

Sarah Johnston is Director of Clinical Trial Transparency at ICON, with over 20 years of industry experience spanning diagnostic assay development, pre-clinical and clinical project management, and technical writing. Over the past 7 years, Sarah has focused on Clinical Trial Transparency within Clinical Research Organisations. Since joining ICON in 2020, she has led diverse, global teams in all aspects of trial transparency, while managing high-profile client relationships and large-scale programs across both pharmaceutical and biotech sectors.

Maureen Kennedy

Maureen Kennedy holds a Bachelor's degree in Interdisciplinary Health Sciences from Saint Joseph's University in Philadelphia and brings over two decades of experience in pharmaceutical research. She joined ICON in 2009 and, for the past 14 years, has specialised in Clinical Trial Transparency. In her current role as Clinical Trial Transparency Manager, Maureen leads a dedicated team focused on the anonymisation and redaction of clinical trial documents for public disclosure.

Michael Healey

Michael Healey is an Associate Director of Clinical Programming at ICON, bringing over 20 years of industry experience. Since joining ICON's Statistical Consulting group in 2021, Michael has specialised in data anonymisation services, submission support, and providing expert guidance on CDISC SDTM and ADaM standards. His career spans a range of technical roles, including database development and analysis programming for regulatory submissions. Additionally, Michael has held managerial positions within ICON's Clinical Programming team, contributing to both technical leadership and team development.