



WHITEPAPER

The Ultimate Guide to Anonymization or Redaction of Documents: Strategies, Workflow, and Mindset



Executive Summary

The demand for transparency in clinical trials is increasing. Pharmaceutical and biotechnology companies – as well as other clinical trial sponsors – face evolving regulatory requirements that necessitate the disclosure of clinical trial data, while sectors such as finance, legal, and government navigate their own privacy obligations. Across all fields, the challenge remains the same: how can organizations disclose information responsibly while safeguarding privacy and protecting sensitive data?

This guide provides a comprehensive approach to anonymization and redaction—equipping teams with the strategies, workflow, and mindset necessary to meet compliance requirements confidently. Drawing on TrialAssure’s proven expertise in this space, this resource offers insights applicable to pharmaceutical sponsors and beyond.



Key Definitions and Distinctions

- **Anonymization:** This involves transforming data so individuals can no longer be identified, directly or indirectly, while keeping its utility for further research.
- **Redaction:** This involves obscuring or removing specific information from documents to prevent disclosure of personal or commercially sensitive details.
- **PII:** Any information that can identify an individual, such as names, dates of birth, or medical history.
- **CCI:** Business-sensitive information, such as trade secrets or proprietary methodologies.
- **Risk-Based Approach:** Effective anonymization and redaction require balancing privacy protection with data utility—ensuring disclosed data retains value without exposing sensitive details.

Understanding when to apply each technique is critical. Pharmaceutical sponsors may anonymize patient data for clinical study reports, while legal professionals might redact confidential client details. Across industries, a clear approach ensures compliance without sacrificing data utility.

An Overview of EMA and Health Canada Regulations

The European Medicines Agency (EMA) and Health Canada have established leading transparency

regulations that require pharmaceutical companies to disclose clinical trial data while safeguarding patient privacy and confidential business information. EMAs Policy 0070 mandates the publication of clinical study reports (CSRs) submitted as part of marketing authorization applications, ensuring that data is accessible to researchers and the public. However, this data must undergo thorough anonymization to protect personal data in compliance with the General Data Protection Regulation (GDPR).

The anonymization process involves removing or modifying identifiers, applying a risk-based approach to balance data utility with privacy protection. Sponsors are required to submit anonymization reports detailing the methods used, ensuring transparency while demonstrating that the risk of re-identification is acceptably low.

Similarly, Health Canada’s Public Release of Clinical Information (PRCI) regulation, introduced in 2019, requires the proactive disclosure of clinical trial data for approved drugs and medical devices. As a result, companies are responsible to redact personal information and commercially confidential information from documents such as CSRs before publication.

Unlike EMA Policy 0070, PRCI provides specific guidance on what constitutes confidential business information and includes a formal reconsideration process if

sponsors believe certain redactions are necessary. Both regulatory frameworks emphasize the need for a structured, reproducible anonymization or redaction strategy, prompting organizations to adopt robust workflows and technology solutions to ensure compliance while preserving data utility.

These changing and increasing requirements force sponsors to adapt quickly, especially when it comes to anonymization. Those who are not able or willing to do so face increased public pressure, the potential of penalties for non-compliance, and issues leveraging advantages that greater transparency can bring.

Greater transparency has the potential to increase:

- Trust, while enhancing public perception
- Awareness and interest from investigational sites and clinical trial participants
- Consistency of information and data released to the public
- The potential reuse of clinical documents and data, thereby utilizing existing research to shorten the drug development process

With greater transparency, ensuring clinical trial participants' privacy and confidential company information is of paramount importance. Strong internal processes around anonymization, redaction of documents, and anonymization of data are needed.



The Three Key Steps to Anonymization

Step #1: Identify and Classify Variables

Before proceeding with the anonymization of clinical information, particularly in documents where the data and information is presented in an unstructured format, it is important to define and classify direct and indirect identifying variables. Direct identifying variables are commonly described as information that meets the following criteria:

- **Replicable:** the variable is unlikely to vary frequently over time
- **Distinguishable:** the individual patients may have distinct, recognizable results or values
- **Knowable:** someone knows a variable, or variables associated with a particular individual

Other identifying variables that fall within the definition of personal information are indirectly identifying variables. These are defined as variables that may present a significant possibility of re-identifying an individual when combined with other available information, like demographic data. Still, these variables may be necessary to understand the clinical data, and therefore their anonymization must be carefully considered and justified.

Variables that do not present a high probability of re-identifying an individual, alone or in combination with other information, are not considered personal information and should not be transformed or redacted during the anonymization process.

Step #2: Measure the Re-Identification Risk

The overall risk of re-identification associated with the disclosure of clinical information is the product of the risk inherent to the data and the risk associated with the release context. While too involved to describe in detail here, a robust quantitative risk assessment process (often paired with a qualitative risk assessment

process) is critical in evaluating re-identification risk. Quantitative risk assessment is typically conducted using a technique called k-anonymity as a basis. This methodology calls for calculating the number of subjects that share common characteristics throughout the data. More advanced techniques, such as l-diversity and t-closeness, build upon k-anonymity with some added sophistication to address possible limitations based on the data distribution. Once the data risk is measured, this risk measurement provides justification for any data transformation(s) that may be employed.

In a public release environment, the risk associated with the context of release is unreducible (i.e., once released, the information is unretractable), so the overall risk of re-identification is higher than the release of information to a small group of select individuals (e.g., researchers). In this scenario, the inherent risk associated with the data becomes the sole driving factor that reduces the overall risk to an acceptable level.

For a successful release of information to the public (i.e., ensuring patient privacy while maximizing data utility), the calculation of re-identification risk needs to reflect this environment. The data itself must be assessed quantitatively so that the overall risk can be measured against a predefined threshold. For example, one must consider the number of subjects that share a common indirect identifier, such as age. If this number is low for any particular value, that potential identifying variable may require anonymization.

Step #3: Anonymize the Data

The methodology used to anonymize clinical information can have a significant impact on data utility. Therefore, it's best not to anonymize variables that do not contribute to the risk of re-identification and to adopt methods that have the lowest impact on data utility.

Directly identifying variables, like name, initials, signature, job title, address, email address, and phone number typically contain little clinical value and should therefore be anonymized through the process of redaction.

Indirectly identifiable variables, like subject/patient assigned ID, city, state, postal code, demographic data, medical history, serious adverse events, dates, height, weight, and BMI, should be considered for transformation or generalization via pseudonymization, so long as the risk of re-identification has been mitigated.

A subsequent risk assessment on the anonymized data consistent with the previous approach is recommended to ensure the shareable data meets acceptable standards and a defined threshold for privacy protection. Adjustments to, or "fine-tuning" of the anonymization rules (redaction and/or pseudonymization) may be needed to ensure maximum data utility while maintaining an acceptable level of re-identification risk.

Who should support these steps?

Transparency subject matter experts (SMEs) are the best suited to conduct anonymization and de-identification of clinical datasets and documents because of their knowledge, background, and expertise in:

- Clinical trial transparency laws, regulations, and guidance
- Clinical trial disclosure, including clinical trial registration and results summary posting
- Management and execution of transparency activities, including policies, protocol, and submissions
- Data privacy and the handling procedures for patient data

Transparency SMEs need to follow good data management and handling procedures such as ensuring an appropriate audit trail for any data transformations is established in anticipation of potential regulatory inspections or sponsor/vendor audits.

They should also seek to use facilitating technology that is efficient, repeatable, and scalable and which

allows them to measure risk and apply changes in an iterative fashion. This also allows for consistency in rules and methodology across the sponsor's pipeline of products.

Additionally, establishing a robust, nimble, scalable, and enduring governance structure and process across all disclosure and transparency-related activities is essential. A responsible data-sharing culture starts with executive sponsorship and should be reinforced across the organization and embedded as part of its mission and vision.



Q&A: Understanding CCI with Zach Weingarden, MS

Anonymizing CCI, and understanding what it is and what it is not, is a hotly-debated topic in the pharmaceutical industry. To increase the understanding of CCI, we speak with Director of Product Solutions Zach Weingarden to answer some of the industry's most anticipated questions.

Question: What is the definition of Company Confidential Information (CCI)? And what are the typical types of information that pharmaceutical companies consider to be CCI?

Answer: Company, or commercially, confidential information (CCI) is information representing a privately held competitive advantage. Or another way to look at it, it is information that could be financially detrimental to

a company if a competitor were to know about it. In the pharmaceutical industry, an example may present itself when a company is in the early stages of investigating using a particular drug to treat a new type of problem. Say this drug is being marketed for another indication, but clinical researchers have an idea based on data or how the drug works as to how it might also be effective for another group of people. The company could use an existing trial to explore this possibility, and if it works, they have a competitive advantage by being the first to do so.

Question: How do pharmaceutical companies determine what information is CCI and what is not?

Answer: This is a complicated question. Mainly, different people have very different ideas about what is, or what should be, confidential. It boils down to whether something is a competitive advantage or not, and that can be extremely hard to define. One clear criterion for CCI is that it must not have been previously shared publicly in any way. I have seen it often where a pharmaceutical executive expects certain information to be considered as CCI, but if it has been shared in a peer-reviewed journal article already, it is considered public information from a legal standpoint.

Question: Regarding regulatory bodies or other government agencies, what safeguards are in place to protect CCI?

Answer: When clinical documents are going to be shared publicly, in Canada or Europe for instance, there is a transparency process where the pharma company is allowed to propose redactions to specific pieces of text that they consider to be CCI. However, the agency will review the underlying text and determine if the redaction is valid based on the company's justifications. Often, the public benefit of sharing the information outweighs the CCI consideration.

Question: How can pharmaceutical companies justify CCI to global health authorities?

Answer: The EMA spells out its general considerations

for CCI pretty clearly. In general, broad information is difficult to justify as CCI. The example I shared earlier about investigating a new indication would likely be rejected. However, specific information about company operations, manufacturing details, or unique data analysis methods which are considered trade secrets might qualify. It is important for companies to understand the agency's position related to the specific type of information so they can align their justification, and expectations, accordingly.

Question: What happens if the CCI is rejected, and what are some common mistakes that companies make when it comes to protecting their CCI?

Answer: The mistake is usually assuming that CCI would be covered in the first place. Around 90 percent of proposed CCI redactions are rejected by the agency. If the proposed redaction is too broad, there might be more specific pieces of information that can be redacted instead, so understanding the agency's policy in advance is helpful. Another common mistake is that the information has already been inadvertently disclosed to the public somehow. Unfortunately, once this has already happened, the cat is out of the bag, and no CCI protections will apply. Companies can take two measures to address this proactively. First, conduct a thorough search of public information before proposing CCI redactions, and second, ensure that all public-facing information releases are reviewed from a CCI perspective to avoid this issue in the first place.

Question: Can information that was once considered CCI become public knowledge over time?

Answer: Of course, the second that CCI is included in any public document it is no longer considered confidential. It could be in a journal article, marketing materials, or even simply discussed in a public forum or presentation – it doesn't matter.

Question: How do companies ensure that employees understand what information is considered CCI and how to protect it?

Answer: There needs to be a unified understanding

across all stakeholders that generate, process, and manage company information of what is CCI and what isn't. This is broader than the transparency department alone. At TrialAssure, we call this a Design for Transparency approach. Everyone in the organization should be educated on this and it shouldn't be up to individuals or functions to make this determination on their own. The overall process of generating and managing information should include steps to track and protect CCI. There must be a mechanism to track and maintain information that is confidential and therefore cannot be shared in any public-facing channel. It must be accessible to all the relevant stakeholders in the organization and of course, kept up to date as information changes or becomes public.

Question: How do pharmaceutical companies balance the need for transparency with the need to protect CCI?

Answer: It's a relatively new challenge since a lot more information is going public than ever before. Many companies are taking the standpoint of assuming that any written document will eventually go public. This is a good thing, but it also means that everyone who is involved in writing these documents should be educated on company specific CCI considerations (and patient privacy too!). This also means that there is a need for intelligent tools to help balance the need to share with the need for privacy and protection of CCI.

Making a Mindset Shift: From Compliance to Confidence

Organizations must move beyond viewing anonymization and redaction as mere regulatory checkboxes. Developing a proactive mindset is key to embedding privacy protection and data disclosure excellence into the core of business operations.

Cross-functional collaboration is the first pillar of this mindset shift. Successful anonymization and redaction

require input from multiple departments—legal, regulatory, clinical, data privacy, and IT teams—to ensure that every angle is considered. This is that “Design for Transparency” approach mentioned earlier. When these teams work together from the outset, they create a cohesive approach that streamlines the process and reduces the risk of errors.

This integrated effort minimizes delays and ensures that privacy safeguards align with both business goals and regulatory standards.

Embracing technology is the second cornerstone of the mindset shift. Automation platforms like [ANONYMIZE®](#) can dramatically increase efficiency and reduce the burden on human reviewers. However, technology alone is not enough. Human expertise remains crucial for interpreting context and making nuanced decisions—especially in complex documents containing both personal and commercially sensitive information.

Organizations that successfully blend automation with human judgment position themselves for both speed and accuracy.

Finally, continuous improvement must become a guiding principle. Regulatory landscapes evolve, and the complexity of data disclosures grows with each passing year. Organizations that treat anonymization and redaction as a static process risk falling behind. Instead, viewing it as an evolving discipline encourages teams to regularly review their practices, adopt new technologies, and stay informed about regulatory updates.

This commitment to improvement fosters a culture of compliance that is both resilient and forward-looking.

Incorporating AI into Anonymization

TrialAssure has harnessed the power of artificial intelligence (AI) to strengthen its anonymization and redaction support. Through advanced machine learning algorithms, TrialAssure ANONYMIZE® has been trained by AI models and industry experts to intelligently

identify and classify personal data and commercially sensitive information within complex documents.

AI-driven risk scoring evaluates the likelihood of re-identification, allowing teams to make data-driven decisions on the appropriate level of anonymization. Additionally, the system can suggest tailored redaction strategies based on document content, improving accuracy and preserving data utility. By integrating AI into the anonymization workflow, TrialAssure enables organizations to achieve regulatory compliance with greater precision and speed, while reducing the burden on human reviewers.



Future Trends and What's Next

Regulatory frameworks around the world are becoming increasingly stringent. The European Union and Canada have led the way with transparency policies requiring the publication of clinical trial data and documents, but all signs point to the United States adopting similar rules. Industry experts anticipate that the FDA will introduce similar anonymization requirements in the near future, driving pharmaceutical sponsors and other industries toward harmonized global practices. Internally, we see how advanced automation and AI-powered redaction will continue to reduce manual burden while improving precision. As standardization emerges, organizations will be better equipped to navigate the intersection of data privacy and public transparency as long as human experts remain in the loop. Throughout all of this, TrialAssure empowers organizations to approach anonymization and redaction with precision and confidence. Whether preparing clinical documents, financial disclosures, or legal records, our solutions ensure that data privacy and compliance go hand in hand.

Learn how TrialAssure ANONYMIZE® can optimize your document workflows.
Email info@trialassure.com for a demonstration or consultation.

