Anonymization of clinical trial data for sharing via Vivli

Introduction

Roche is committed to delivering on the principles of clinical trial transparency as set out in the joint EFPIA-PhRMA Principles for Responsible Clinical Trial Data Sharing including sharing patient level data in compliance with global legal and regulatory requirements.

Roche is committed to sharing clinical trial data in a transparent manner while maintaining the privacy and confidentiality of the research participants. Roche shares clinical data because we understand it helps physicians, patients and healthcare providers to make informed treatment decisions. Data Sharing can also enable researchers to more easily build on our research and the research of others, in the hope of advancing scientific progress. In past years, we have taken steps to advance how we share our clinical research, both within the scientific community and the broader public.

This document describes the approach taken by Roche to prepare data for sharing on the Vivli platform. These steps are taken to minimize the risks to the privacy and confidentiality of the research participants and to ensure compliance with data privacy legal requirements. Roche's approach to data sharing is governed by Roche's data sharing policy link.

General Approach

Approved data requests will be granted access to the relevant anonymized data packages for each study. The anonymization packages contain patient level data and accompanying documentation that help give context to the data, including clinical study reports and study protocols. Each dataset and document has to go through the Roche anonymization process, this document describes how Roche processes those datasets.

Anonymization Framework

Roche has adopted a practical and risk-based anonymization framework which aims to reduce the risk of re-identification in a manner that is proportional to the data and context risk. The overall aim is for a pragmatic balance; ensuring an acceptably low risk of re-identification whilst retaining data utility. The assessment of data risk involves looking at what potential identifiers exist in the data, the sensitivity of the dataset and the impact of re-identification on data subjects. The assessment of context risk includes looking at what technical and organizational measures are in place to control the sharing (e.g. governance, IT system access controls, processes etc.), how the data will be shared, as well as the potential re-identification attack scenarios and adversaries. By balancing the extent to which datasets are obfuscated with the risk of sharing in the context of the Vivli platform allows a higher level of data utility to be maintained.

The overall risk assessment framework we have adopted is based upon the concepts from the Guide to the De-Identification of Personal Health Information (El Emam, K. (2013). Boca Raton,

FL: CRC Press.) and the Anonymization Decision-making Framework (Elliot, M., Mackey, E., O'Hara, K., Tudor, C. (2016). Manchester, UK: University of Manchester). The methodologies developed by these groups are highly referenced and well accepted across privacy experts. The risk assessment frameworks slightly differ, however, similar key process steps have been adopted including: determine potential direct and indirect identifiers in the data; identify possible 'adversaries' and plausible attacks on the data; consider data utility; determine the risk of re-identification threshold and evaluate risk of re-identification (qualitatively); apply anonymization methodology and document the anonymization methodology and process.

Together with the risk assessment frameworks, we have also adopted pharmaceutical industry best practices, including PHUSE de-identification standards (PHUSE De-Identification Working Group. (2015). De-Identification Standard for CDISC SDTM 3.2), and Transcelerate standards (Transcelerate Biopharma. (2016). De-identification and Anonymization of Individual Patient Data in Clinical Studies). PHUSE and Transcelerate are cross-pharma forums and represent the latest industry thinking and approaches.

Anonymization Methodology

The starting point for the anonymization process is 'pseudonymized' clinical trial data (that is personal data labelled with a pseudonym). The anonymization methodology involves one-way hashing of patient, site and sample identifiers, obfuscating or re-coding other potential identifiers, offsetting dates, and generalizing data such as geographical locations, as well as other changes specific to the data. Unlike direct identifiers, many indirect identifiers are important for data utility and if removed completely may limit the ability of researchers to perform complete analyses (or re-create original analyses). Following a risk-based anonymization methodology allows removing or modifying only the indirect identifiers which pose a risk of re-identification while also maintaining data utility. Overall, the Roche anonymization approach ensures that the risk of re-identification is low across Vivli and other data sharing scenarios.

The methodology is implemented by an internal specialist group, expert in anonymization approaches. The anonymization approaches for external data sharing have been used in practice since 2014 and are regularly being reviewed and adjusted in particular considering technological developments. In that time we have shared clinical trial data over 600 times (some studies being shared multiple times), including via Vivli. There have been no known re-identification attacks or successful re-identifications, which would indicate the robustness of the approach.