# Takeda Clinical Trial Data Transparency
# Anonymization and Data Protection Procedures

## Introduction

Protecting the privacy of patients or study participants (hereinafter referred to as participants) who consent to contribute their personal data collected as part of a clinical trial investigation is an important obligation of sponsors conducting clinical trials. As a sponsor organization, Takeda Pharmaceutical Incorporated Limited takes appropriate measures to ensure that participant privacy is safeguarded through various methodologies including anonymization of their data.

This document describes the general approach taken by Takeda to prepare participant-level clinical trial datasets for the purposes of external data sharing with qualified researchers in response to requests for such information received via www.Vivli.org. This approach minimizes the risks to the privacy and confidentiality of study participants, while ensuring compliance with legal requirements pertaining to data privacy. Takeda seeks to comply with all applicable laws and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and European Union regulations on processing and protection of personal data including the General Data Protection Regulation (EU) 2016/679 (GDPR). In line with these regulations, for each study in a data sharing project there is a risk assessment for anonymization approval.

## General Principles

a. For each approved data sharing project, Takeda will maintain the integrity between the datasets to allow reliable analyses.

b. Takeda will share the anonymized participant-level data in the same structure that was used for the original CSR analysis, unless otherwise specified.

   **Note**: Only data submitted to an agency will be provided for anonymization.

c. The extent of anonymization will be modified based on multiple risk factors, such as the study population, disease prevalence, data sensitivity, system controls, and others.

## Process Overview

The following steps will be performed when anonymizing clinical trial data for external research in an anonymous format:

1. Removal of personally identifiable information (PII) from each dataset.
2. Internal quality control (QC) review and approval of the anonymized dataset.
3. Destruction of intermediate anonymized outputs and storage of the anonymized dataset.

**Takeda Clinical Trial Data Transparency**

**Anonymization and Data Protection Procedures**

## 1. Removal of PII from each Dataset

Personally identifiable information (PII) is defined as any information relating to an identified or identifiable natural person (ie, anyone who can be identified directly or indirectly by reference to released PII, such as identification numbers or factors specific to that person's physical, physiological, mental, economic, cultural, or social identity). This data is sometimes also referred to as personal-private information (PPI) or protected personal data (PPD).

Takeda's methodology to remove PII from the datasets include:

a. Original participant and site numbers will be replaced with pseudo-participant and pseudo-site numbers.
   i. Each unique participant number is replaced with a corresponding unique randomly generated pseudo-participant number.
   ii. Similarly, each unique site number is replaced with a corresponding unique randomly generated pseudo-site number.
   iii. The same new participant and site numbers are used for the trial to enable participant data to remain linked.
b. Participant initials are cleared, if they exist.
c. Participants' age is calculated, if needed, from the corresponding birth date prior to being cleared.
   i. Age is aggregated when distribution of values is considered insufficient.
   ii. For participants >89 years old, these participants are aggregated into a single category of ">89".
d. All original dates (other than birth dates) relating to a participant are replaced with pseudo dates. A random offset in number of days is generated for each participant and added to all dates for that participant.
   i. All original dates are then replaced by the new dates so that the relative days for each participant are retained.
   ii. Any partial dates are truncated to only the year.
   iii. Imputed dates, in most cases, are made blank.
e. Investigators' identification numbers and names are cleared in all datasets.
f. All comment, reason, and specify fields from all datasets are cleared. Comments datasets such as SDTM CO will not be provided.
g. Values for variables containing free text verbatim terms are cleared in the dataset, including Adverse Events, Medications, Medical History, and any other datasets that have such verbatim text variables.

h. Regarding coded datasets, only 1 level of dictionary term is kept for each record unless the only term available is the verbatim.
i. All other character variables in the datasets are reviewed. Values for any variable found to contain direct or indirect personal identifiers for any participant will be cleared for all participants.
j. Numeric values that could contain PII, such as height and weight, are reviewed and cleared if needed.

## 2. Internal QC Review and Approval of the Anonymized Dataset

Takeda verifies each anonymized dataset through QC, to confirm that no PII remains, as follows:

a. A Data Integrity Check is conducted to compare the anonymized database to the source data verifying:
    1. The number of records matches for all datasets.
    2. The applicable changes listed in this document were appropriately applied.
    3. No unplanned changes were inadvertently made.
b. Documentation of the verification steps, status, and any approvals.

## 3. Project Clean up and Storage of the Anonymized Datasets

Once QC work on anonymized datasets has been completed, the following will occur if they are available to Takeda:

- Key dataset(s) (datasets containing the links between original values and new values in the anonymized datasets, including participant and site numbers) will be permanently deleted.
- Any seeds utilized for random number generation for replacing original values are blanked out.
- Any QC output datasets containing PII are deleted.
- Any Log or Listing files containing PII are deleted.

The anonymized datasets are stored in a secure location separate from the original datasets.