Takeda Clinical Trial Data Transparency
Anonymization and Data Protection Procedures

Introduction

Protecting the privacy of subjects or study participants (hereinafter referred to as participants) who consent to contribute their personal data collected as part of a clinical study investigation is an important obligation of sponsors conducting clinical studies. 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 study 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 study, 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 statistical analysis, unless otherwise specified.

c. The extent of anonymization will be modified based on multiple 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 study 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. Project clean up and storage of the anonymized datasets.

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 is shown in a-j below; see also Section 4 of this document for a succinct Summary of Anonymization Standards:

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 study 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 imputed to full dates, shifted by the date offset, then 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. Adverse Events, Concomitant Medications, Medical History datasets: if the dataset is coded then one dictionary term is kept and other dictionary terms and verbatim event text variables are cleared. If there are no dictionary terms then at least one event variable will be shared (with PII removed from individual values)

g. Comment, narration, reason, specify, and verbatim text variables are cleared.
h. Comments datasets such as SDTM CO will not be provided. Other datasets may be excluded based on sensitivity or lack of data utility.
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:
   i. The number of records matches for all datasets.
   ii. The applicable changes listed in this document were appropriately applied.
   iii. 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 be deleted if they are available to Takeda:

- Key dataset(s) (the dataset containing the links between original values and new values in the anonymized datasets, including participant and site numbers)
- Seeds utilized for random number generation for replacing original values
- QC output datasets containing PII
- Log or Listing files containing PII

The anonymized datasets are securely stored.

Takeda Clinical Trial Data Transparency

Anonymization and Data Protection Procedures

4. Summary of Anonymization Standards

The following data protection procedures for personally identifiable information (PII) are performed by Takeda prior to data sharing. Note that some procedures require data-dependent decisions made on a per study basis, predominately based upon the number of subjects enrolled into the study, as required to protect subject privacy, resulting in different actions taken from study to study.

Data Point	Action Taken	Clarification
Subject identifiers	Replace	by new subject number, keeping the subject-record relationship
Site number	Replace	by new site number, keeping the subject-site relationship
Other direct subject identifiers	Suppress	e.g. kit number, device id
Other indirect subject identifiers (Including: age, race, sex, ethnicity, weight, height, BMI)	Retain	if not identifying
	Aggregate into bands	if low frequencies
	Suppress	outlier or low frequency values
Country	Retain	if one country
	Aggregate to Region	if more than one country
Adverse events, medical history, concomitant medications	Retain	only the lowest level decoded dictionary term. Note: if no dictionary terms then at least one event variable will be retained and PII suppressed
Comments and verbatim terms	Suppress	Note: comments datasets such as SDTM CO will not be provided
Other variables containing PII	Suppress	e.g. investigator names, lab names
Date of birth	Suppress	Note: age will be retained
Subject study dates (full)	Shift by date offset	Apply a randomly chosen offset value for each subject
Subject study dates (partial)	Retain year only	Impute to a full date, shift by the date offset, then share year only
Subject study dates (imputed)	Suppress	
Subject study times and study days	Retain	