Takeda Clinical Trial Data Transparency
Anonymization and Data Protection Procedures

Introduction

Protecting the privacy of patients who participate in clinical trials is an important obligation of sponsors who conduct clinical trials. Takeda will take appropriate measures, including anonymization of data, to ensure that patient privacy is safeguarded. This document describes the approach taken by Takeda to prepare patient-level (hereinafter referred to as "subject") clinical trial datasets for sharing with qualified external researchers in response to requests for such information via www.clinicalstudydatarequest.com. This approach minimizes the risks to the privacy and confidentiality of research subjects and ensures compliance with data privacy legal requirements.[1-4]

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 subject-level data in the same format that was used for the original statistical analysis.

Process Overview

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

a) Removal of personally identifiable information (PII) from each dataset.[1-3]

b) Quality Control (QC) review and approval of the anonymized data to confirm that no PII remains.

c) Secure storage of the anonymized dataset in a designated location.

d) Destruction of the code-key which links the anonymized dataset and the original dataset.

1. Details of Removing PII from the Datasets

a) Original subject and site numbers will be replaced with pseudo-subject and pseudo-site numbers. Each unique subject number is replaced with a corresponding unique randomly generated pseudo-subject number. Similarly, each unique site number is replaced with a corresponding unique randomly generated pseudo-site number. The same new subject and site numbers are used for the trial to enable subject data to remain linked.

b) Subject initials are cleared.

c) Subjects' birth dates are cleared. For subjects >89 years old, the numeric value of age is set to blank, and these subjects are aggregated into a single category of "90 or older," unless an existing age group already includes 90+ (i.e. > 75 years old). For subjects ≤89 years old, age is retained.

d) All original dates (other than birth dates) and times relating to a subject are replaced with pseudo dates and times. A random offset in number of days is generated for each subject and added to all dates for that subject. All original dates are then replaced by the new dates so that the relative days for each subject are retained. Similarly, a random offset in number of minutes is generated for subject and applied to all times for that subject. All original times are then replaced by the new times so that the relative time between any of a subject's records is retained.

e) Investigators' identification numbers, names, and locations are cleared in all data sets.

f) All comment fields from all data sets 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) All dictionary coded terms that use a finite pre-specified list are retained.

i) All other character variables in the datasets are manually reviewed. Values for any variable found to contain direct or indirect personal identifiers for any subject will be cleared for all subjects.

2. Internal QC Review

A reviewer will check each anonymized dataset as follows:

a) Data Integrity Check: Compare the anonymized database to the source data to verify the following:

1. Number of records match for all datasets
2. All the applicable changes listed in this document were applied.
3. No fields were altered except as listed in this document.

b) Compare key summary statistics generated using the anonymized datasets against reports previously submitted from the source datasets. This is to verify that previously reported results are reproducible for variables containing PII.

3. Destruction of the Code-key, Intermediate Anonymized Outputs, and Storage of the Anonymized Datasets

Once QC work on anonymized datasets has been completed, destruction of the following will occur:

- Code-key dataset (the dataset containing the links between original values and new values in the anonymized datasets, including subject and site numbers).

- The seeds utilized for random number generation for replacing original values.
- Any QC output datasets containing PII.
- Any Log or Listing files containing PII.
- Any intermediate copies of anonymized datasets.

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

References

1. (HIPAA) CFR – Title 45: Public Welfare, Subtitle A, 164.415
2. Regulation (EC) No. 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data.
3. Directive 95/46/EC of the European Parliament and of the Council of October 24, 1995 on the protection of individuals with regards to the processing of personal data and on the free movement of such data.
4. Hrynaszkiewicz I, Norton ML, et al. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ 2010;340: c181.