Anonymization standard: Taiho Policy for Anonymization of Clinical Trial

1. Introduction

In line with TAIHO PHARMACEUTICAL CO., LTD., on behalf of itself and its subsidiaries (“Taiho”), corporate philosophy — “We strive to improve human health and contribute to a society enriched by smiles”— Taiho as a company strives to keep providing patients worldwide with innovative drugs.

Taiho recognizes that access to clinical trial data is valuable for the public health, data transparency and science. Taiho has also an obligation to protect the privacy and confidentiality of research participants’ personal information while providing access to clinical trial data to allow for further research.

The objective of this document is to describe the high level approach that is required to perform anonymization of independent patient data in response to researcher’s requests.

Specific privacy laws and regulatory guidance documents (e.g., EU data protection requirements and the US Code of Federal Regulations - Title 45: Public Welfare, Subtitle A §164.514) must be followed as part of this process, and are reflected in the approach for data anonymization and protection described in this document.

This document describes the approach taken by Taiho to prepare data for sharing with other researchers in a way that:

☐Minimizes risks to the privacy and confidentiality of study subject.
☐Ensures compliance with applicable data privacy requirements.

2. Anonymization Approach

2.1 Overview

Taiho will provide anonymized data for analysis to researchers based on approved research proposals. Anonymization involves:

a. **Removing personally identifiable information (PII) from the dataset.**

This includes:

- Recoding identifiers
- Removing comments, free text and free text verbatim terms (“Free Text”)
- Replacing date of birth with age at study entry (or age category, when applicable),
- Replacing all original dates relating to individual subjects with randomly generated offsets which are then applied to create ‘dummy dates’ or replacing them with a ‘study day’.

b. **Ensuring there is no direct link between the anonymized dataset and the original dataset.**

Data can only be considered anonymized if personally identifiable information is removed (or redacted) and the new code number cannot be linked to a specific research subject.

2.2 Specific approach to remove personally identifiable information from the dataset

The 18 identifiers (as defined by HIPAA –see U.S. Code of Federal Regulations - Title 45: Public Welfare, Subtitle A §164.514) are removed from the datasets and any other PII presented below are also removed for dataset anonymization.

Other information that could result in subject identification will be evaluated for removal or recoding, which may include:

- Any names (persons or institutions/companies) and initials,
- Kit numbers and device numbers,
- De-identifying geographic information that applies to specific subject groups (e.g. zip code, place of work). In some cases, country can be removed and replaced with continent.
- Socioeconomic data (occupation, income or education, household and family composition and multiple pregnancies)

In addition the following steps are undertaken:

- Recoding identifiers.
- Removing free text verbatim terms.
- Replacing date of birth with age at study entry.
  Ages above 89 which are aggregated into a single category of “90 or older.”
- Replacing all original dates relating to individual subjects with Study Day.
- Reviewing and removing other PII

These steps are described in further detail below.

2.2.1 Recoding Identifiers

- Research subjects’ identification code numbers (and any other code keys used in the study data, if applicable) are anonymized by replacing the original code

number with a new code number (the “New Identifier(s)”) and destroying the link between the two code numbers.

• The New Identifiers are used across all datasets applicable to a single study (e.g., raw dataset, analysis-ready dataset)

• The investigator identifier (or code number) is re-coded for each investigator, and instigator name are removed.

• Site identification information is re-coded or set to blank.

2.2.2 Removing Free Text verbatim terms.

Information contained in Free Text may compromise a subject’s anonymity. Therefore, removal of comment field text and most Free Text will be performed. Specifically:

- Free Text verbatim terms* are set to “blank” or removed from the dataset including:
  - Adverse Events
  - Medications
  - Medical History
  - Other specific verbatim free text

  *The standard terminology (typically Adverse Events, Medications, and Medical History) coded from free text will be provided.

2.2.3 Replacing Date of Birth

Information relating to a research subject’s date of birth and identification of specific ages above 89 may compromise a subject’s anonymity. Date of birth is replaced with age at study entry with the exception of ages above 89, which are aggregated into a single category of “90 or older.”

2.2.4 Replacing all Original Dates relating to a Study Subject

Information relating to all Original Dates may compromise a subject's anonymity. Study sponsors use one of two methods as described below.

2.2.4.1 Dummy Date Method

All original dates are replaced from datasets: A random offset is generated for each research participant and applied to all original dates for that research participant. All original dates are replaced with the new dummy dates so that the relative times for each research participant are retained.

Example: If the original reference date was 01APR2008 and the date of death was 01MAY2008, a random offset is generated (in this case 91 days). Dummy dates are than calculated using this offset of 91 days.

	Original Date	New date	
Reference Death	01Apr2008	01Jul2008	Apply offset = 91 days
Date of Death	01May2008	31Jul2008	Apply offset = 91 days
Relative Time of Death	30 days	30 days	

2.2.4.2 Study day method

All original dates are removed from the datasets. The Study Day is calculated for each observation with days relative to a reference date. In order of priority, the reference date is defined as the date of first study treatment, date of randomization, or date of consent. For example if a subject is randomized, but does not take the study treatment (i.e., the date of first treatment is missing), the date of randomization will be used as the reference date to calculate the study day for any assessments recorded.

Example: If the original reference date was 01JAN2008 and the date of death was 01MAY2008, the date of death would be 122 as the calculated Study Day.

	Original Date	Reference Date	Study Day
Date of Death	01May2008	01Jan2008	122

2.2.5 Reviewing and Removing/Redacting Other PII

Other data elements that contain PII are removed. For example:

• Information from variable names (e.g. lab names may contain location information)

• Investigator comments that may be used to identify a subject

• Genetic data that may enable a direct trace back to an individual subject.

• Supplemental data (e.g., case narratives, documentation for adjudication, imaging data (X-rays, MRI scans), etc.)

2.3 Review and Quality Control

Quality control checks are performed and documented following the creation of the anonymized data and supporting documents. The anonymized data and supporting documents are stored in a separate secure location from the original study datasets.

2.4 Destroying the link (key code) between the dataset that is anonymized and the original dataset

Research subjects' identification code numbers are anonymized by replacing the original code number with a new code number (see section 2.2.1), and destroying the code key that was used to generate the New Identifier from the original identifier.

The following specific items are discarded:

- Any transactional copies of anonymized datasets
- De-identification tables (links for original variable and new

anonymized variable)

- Any QC output datasets
- Any Log or LST files
- The seed utilized for random number generation

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

3.0 Exceptions

Any study, where the data set cannot be anonymized to ensure subject privacy, according to Taiho assessment, will not be transferred to researchers.