Otsuka Procedures for Clinical Trial Data Anonymization

1. Introduction

Otsuka Pharmaceutical Co., Ltd., on behalf of itself and its subsidiaries ("Otsuka"), as a company focused on innovation, recognizes that access to clinical trial data is valuable for the advancement of public health and science. Otsuka has 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 all trial participant ("subject") level 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 Otsuka to prepare data for sharing with other researchers in a way that:

- Minimizes risks to the privacy and confidentiality of research subjects.
- Ensures compliance with applicable data privacy requirements.

2. Anonymization Approach

2.1 Overview

Otsuka 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 such things as recoding identifiers (by replacing the original code number with a new code number), 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) and replacing all dates relating to individual subjects with "Study Day" (see section 2.2.4 below).
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 related documentation as the initial step for dataset anonymization.

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

- Any names and initials,
- Kit numbers and device numbers,
- Any geographic information more specific than country, such as zip code or place of work,
- Socioeconomic data such as occupation, income or education, household and family composition and multiple pregnancies.

In addition the following steps are undertaken:

- Recoding identifiers,
- Removing Free Text,
- 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 code key (or algorithm) that was used to generate the new code number from the original (i.e., 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).

Extension studies use the New Identifiers as well, to enable individual subject data to remain linked. This also applies to long term follow-up studies where separate reports are published. This is achieved by repeating the data anonymization process for the initial study data at the same time as the extension/follow up study data.

2.2.2 Removing Free Text

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. In most situations, these Free Text fields are coded to standard terminology and these coded terms will be provided. Free Text variables that will be removed include:
  - Adverse Events
  - Medications
  - Other specific verbatim free text
- All dictionary coded terms (typically Adverse Events, Medications, and Medical History) with decoded and/or terms that use a pre-specified list are retained.
- Additional study specific or subject specific data may be considered in the anonymization process and removed from the study data.
- Certain Free Text Fields may be retained if they do not contain PII and removal of these fields may impact the scientific value of the dataset (e.g. medical history that has not been coded).

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 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

All 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 Removing location identifiers and aggregating study site information

All variables containing investigator and study site specific information (e.g., investigator name and city name of site) will be removed and set to 'blank' or replaced with a new code to prevent identification of location of a subject. Countries with limited study sites and/or subjects may be aggregated or assigned with a new code so that location cannot be identified. Studies with limited study sites or subjects may not be made available to researchers due to risk of re-identification.

2.2.6 Exclusion of supplemental data and other PII

Supplemental data (e.g., case narratives, documentation for adjudication, imaging data (X-rays, MRI scans), etc.) will NOT be made available. Other data elements that contain PII may be removed, such as:

- Information from variable names (e.g., lab names may contain location information)
- Genetic data that would enable a direct trace back to an individual subject

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, or New Identifier (see section 2.2.1 above), and destroying the code key that was used to generate the New Identifier from the original identifier (i.e., destroying the link between the two code numbers).

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 Clinical Study Report

For approved requests, the Clinical Study Report document, limited to the report body and appendices, will undergo a redaction process to remove subject PII. A similar set of items as those listed in the sections above will be redacted from the Clinical Study Report.

4.0 Exceptions

Any study, where the data set cannot be anonymized to ensure subject privacy, according to Otsuka assessment, will not be transferred to researchers. In such cases, a written justification will be completed, and will be available upon request.