

# **Boehringer Ingelheim Clinical Trial Data Transparency Anonymisation and Data Protection Procedures**

The objective of this document is to describe the high level steps that are required to perform anonymisation of all subject level data (“data”) prior to export into the Data Access System (DAS).

The following list summarises all anonymisation steps performed by Boehringer Ingelheim (BI) prior to data transfer into the DAS.

## **A. Removing personally identifiable information (PII)**

The 18 identifiers (as defined by HIPAA – see US Code of Federal Regulations - Title 45: Public Welfare, Subtitle A §164.514) – where recorded - and other personal identifiers that may be present are removed from the datasets and related documentation.

This involves removing:

- any names and initials,
- kit numbers and device numbers,
- geographic information lower than country level,
- information from variable names e.g. lab names may contain location information.

In addition the following steps are undertaken:

- A.1 Recoding identifiers and formal anonymisation
- A.2 Replacing date of birth
- A.3 Aggregating centre information to the country/region level
- A.4 Replacing original dates related to a study subject
- A.5 Removing comments, free text and free text verbatim terms

These steps are described in further detail below.

### **A.1 Recoding identifiers and formal anonymisation**

Some Data Protection Authorities in Europe suggest that the data can only be considered anonymised (for research purposes) if personal information is removed (or redacted) and the subject code number cannot be linked to a study participant.

Therefore, study subject numbers are anonymised by replacing the original code number with a new code number and destroying the code key that was used to generate the new code number from the original (i.e. destroying the link between the two code numbers). Any other code keys used in the study data will be deleted or recoded with the code key destroyed.

## **A.2 Replacing date of birth**

Date of birth is replaced with age in years and all ages above 89 are aggregated into a single category of “90 or older”. This is a specific HIPAA requirement

## **A.3. Aggregating centre information to the country/region level**

All variables containing centre and/or investigator specific information (such as the investigator name and the city name of the site) will be removed or set to ‘blank’ to prevent identification of the location of a subject.

Countries with only one centre will be aggregated to a region, so that location cannot be identified. For the same reason, studies with one centre cannot be anonymised and will not be made available to the DAS.

For avoidance of doubt, the resulting lowest geographical subdivision of the data is country. Therefore, the BI data provided to DAS is stricter than the geographical anonymisation standards listed as data element 2 in the Health Insurance Portability and Accountability Act of 1996 (HIPAA).

## **A.4. Replacing original dates related to a study subject**

The replacement of dates prevents subject identification in case a date is associated with a critical event. The original dates will be shifted by a random, subject specific factor. Therefore, differences between dates within the subject’s data records will be maintained.

In cases where only the year of an event is collected, this will not be modified. Where available, durations, e.g. time to event information, will be retained.

## **A.5. Removing comments, free text and free text verbatim terms**

Removal of comment field text and of any free text is essential; as such fields may contain very patient-specific information and therefore may allow identification of a data subject. All comment fields and all free text will be removed. This also includes the verbatim terms on the Case Report Forms (CRFs) for reporting of Adverse Event (AE), Concomitant Diagnoses and for Concomitant Therapy. Access to the verbatim information from these three CRFs will be available for structured analysis, as the corresponding coded medical dictionary terms are included in the datasets.

## **B. Process and Quality Control**

The annotated CRF and the dataset/variable content of the raw datasets and the analysis-ready datasets (used for BI’s analysis) are reviewed to identify those variables to be processed according to the rules described above. This manual review is further supported by automated indicators, in particular to identify PII.

Quality control checks are conducted for the processing of the data and supportive metadata documentation. This includes verification that the code key was destroyed.

**C. Exclusion of supplemental data from data sharing**

Supplemental data, e.g. case narratives, documentation for adjudication, imaging data (X-rays, MRI scans, etc.) will not be made available to the DAS.

**D. Exclusion of studies which are not possible to anonymise**

Clinical studies of rare diseases and single centre studies must not be made available to the DAS because for these anonymisation is not achievable with reasonable effort and in accordance with applicable data protection requirements. In addition, there may also be studies involving high risk sensitive data (e.g. studies related to specific mental or sexual diseases). For such studies BI will assess the feasibility of anonymisation as part of the review of enquiries, and, in case of a negative outcome, will not provide patient level data but try to address requests by providing summary data or otherwise.

**Summary**

**The result of the BI anonymisation process described above is a data set, which is considered an anonymised subject data set.**

**Any study, where the data set cannot be anonymised, according to BI assessment, will not be transferred to the DAS. Cases of doubt will be forwarded to the BI Data Protection officer for consultancy. The final decision on anonymisation and transfer to the DAS will be made by the BI Data Sharing Request Review Committee (DSRC). If the DSRC decides that a dataset cannot be anonymised, a written justification will be provided and will be publically available on request.**