DATA PROTECTION PROCEDURES FOR INDIVIDUAL PATIENT DATA

The following list summarises all data protection procedures performed by Boehringer Ingelheim (BI) prior to data sharing.

1 REMOVING PERSONALLY IDENTIFIABLE INFORMATION (PII)

The 18 identifiers (as defined by the Health Insurance Portability and Accountability Act of 1996 (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 procedures are undertaken:

1.1 Recoding identifiers
1.2 Replacing date of birth
1.3 Removing site information and aggregating country to the region level
1.4 Replacing original dates related to a study subject
1.5 Removing comments, free text and free text verbatim terms

These procedures are described in further detail below.

1.1 RECODING IDENTIFIERS

Study subject numbers are anonymised by replacing the original code number with a new code number. Any other code key used in the study data will be deleted if it is not used to link the dataset with other datasets. If such a code key is used to link the dataset with other datasets, it will be recoded.

1.2 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.

1.3 REMOVING SITE INFORMATION AND AGGREGATING COUNTRY TO THE REGION LEVEL

All variables containing site 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 re-identification of the location of a subject.

For a clinical trial with physical sites, countries with only one site will be aggregated to the region level, to prevent re-identification of the location of a subject. For the same reason, a clinical trial with one physical site cannot be anonymised and will not be shared.

For a Decentralized Clinical Trial conducted with virtual sites, there is no need to aggregate country to the region level.

The resulting lowest geographical subdivision of the data is country. Therefore, the BI requirements are stricter than the geographical anonymisation standards listed as data element 2 in the HIPAA.

1.4 REPLACING ORIGINAL DATES RELATED TO A STUDY SUBJECT

The replacement of dates prevents subject re-identification if a date is associated with a critical event. The original date values will be shifted by a subject specific random factor, i.e., all dates of a subject are shifted by the same factor.

In cases where only the year of an event is collected, the original year values will be shifted by the subject specific random date shift factor. Where available, durations, e.g., time to event information, will be retained.

1.5 REMOVING COMMENTS, FREE TEXT AND FREE TEXT VERBATIM TERMS

Removal of comments and of any free text is essential. Such fields may contain patient-specific information and therefore may allow re-identification of a study subject. Also, the free text verbatim terms from the Adverse Events (AE), Medical History (MH) and Concomitant Medications (CM) domains and sub-domains will be removed to avoid re-identification of a data subject. For AE and MH the corresponding coded dictionary terms will be added to the datasets for analysis. For CM no replacement by coded dictionary terms will be performed by default due to license restrictions.

2 PROCESS AND QUALITY CONTROL

The annotated Case Report Form (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, 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.

3 EXCLUSION OF SUPPLEMENTAL DATA FROM DATA SHARING

Supplemental data, e.g., case narratives, documentation for adjudication will not be shared.

4 EXCLUSION OF STUDIES WHICH ARE NOT POSSIBLE TO ANONYMISE

Clinical studies of rare diseases and single site studies must not be shared 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 dataset, which is considered an anonymised subject dataset.

Studies where the dataset cannot be anonymised according to BI assessment will not be shared. Cases of doubt will be forwarded to the BI Data Protection officer for consultancy. If a dataset cannot be shared, a written justification will be provided and will be made publicly available on request.