DATA PROTECTION PROCEDURES FOR INDIVIDUAL PATIENT DATA

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

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 procedures 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 procedures are described in further detail below.

A.1 Recoding identifiers and formal anonymisation

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 shared.

For avoidance of doubt, the resulting lowest geographical subdivision of the data is country. Therefore, the BI data shared 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, the year will be shifted by the random, subject specific date shift factor. 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 Form (CRF) for reporting of Adverse Events (AE), Concomitant Diagnoses and Concomitant Therapies. Access to the verbatim information from these three CRF pages 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 shared.

D. Exclusion of studies which are not possible to anonymise

Clinical studies of rare diseases and single centre 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 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 shared. Cases of doubt will be forwarded to the BI Data Protection officer for consultancy. If a data set cannot be shared, a written justification will be provided and will be made publically available on request.