EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Work instructions

Title: Design of an analysis of EudraVigilance data		
Applies to: Pharmacovigilance and Epidemiology Department		
Status: PUBLIC		Document no.: WIN/H/3290
Lead Author	Approver	Effective Date: 17-DEC-18
Name: Luis Pinheiro	Name: Peter Arlett	Review Date: 17-DEC-21
Signature: ON FILE	Signature: ON FILE	Supersedes: WIN/H/3290 (20-AUG-10)
Date: 17-DEC-18	Date: 17-DEC-18	TrackWise record no.: 5449

1. Changes since last revision

Updated to reflect changes made to SOP/H/3289.

2. Records

Analyses of EudraVigilance (EV) data should be stored under the relevant product folder in DREAM under a newly created folder in Cabinets/13. Programmes and Projects/zz. Closed projects 2004-2014/EudraVigilance - NEW STRUCTURE/Pharmacovigilance/Data Analysis/DWH Analysis/Product Analysis/<corresponding year>/<subdirectory>

3. Instructions

3.1. Data analysis plan

The data analysis team (DAT) should elaborate a data analysis plan (DAP). The DAP should be based on the results of the exploratory data analysis and describe the purpose and the methodological options that will be used to analyze the data in EV.

It should be a short synopsis and include basic sections, namely:

- Roles and responsibilities;
- Milestones and proposed timelines;

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 Facsimile +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

- Purpose or objective of the analysis;
- Specification for data extraction;
- Case definition;
- Analysis technique(s) to be applied;

3.1.1. Roles and responsibilities

The roles and responsibilities section should detail the members of the data analysis team and their roles, as well as the peer reviewers and sign-off person(s).

3.1.2. Milestones and proposed timelines

The milestones and proposed timelines should consider the main milestones of the analysis and the expected time to complete those. This is particularly relevant if additional data cleaning is required.

3.1.3. Purpose or objective of the analysis

The purpose or objective of the analysis should match request or the scope of the procedure, if one has been initiated or is ongoing.

If a request provides detailed objectives for the analysis, (e.g. perform a disproportionality analysis restricted to a class of products) but the DAT considers that there is a more useful alternative to address the concern, this should be clearly highlighted in the DAP and agreed with the requestor.

3.1.4. Specification for data extraction

The analysis should be run on raw data extracted at a given point in time. This is due to two reasons: 1) traceability, as results in EV can change over time due to nullifications, recoding of products, updates of MedDRA terms in follow-ups, etc. and 2) reproducibility, as having the raw data allows conducting repeat, revised or new analyses on the same data.

The data extraction criteria should be clearly defined in the DAP. Typically the product identification will be straightforward, however some data analyses requests may describe the medicinal products involved in either a general, class-level terms (e.g. quinolones) or be extremely selective , e.g. “cyproterone acetate / ethinylestradiol (2 mg/0.035 mg) containing medicinal products”.

The appropriate data extraction criteria should be defined to allow for the correct collection of data. This may need to involve the requestor and procedure lead, if there is one.

Where a balance between precision, i.e. ensuring that only true cases are extracted at the cost of missing some true cases, and recall, i.e. ensuring that all true cases are extracted at the cost of including false positive cases, is required, the DAT should give priority to recall.

Particular attention should also be given to the querying the data using time. Receive, receipt, gateway and reaction start dates are all different. Querying cases with EV gateway date will produce different results to querying cases with Receive date. Receive date is the closest proxy to detection date, thus for time series, this should be the field used.

The need for additional data quality management should be considered at the exploratory data analysis stage following the request to perform an analysis of EV data. Reasonable efforts should be made by the DAT to confirm that no duplicate is present in the extracted dataset and that the data set is of the highest possible quality.

The filters that are applied on ICH E2B R2/R3 fields should be described as such, e.g. D.5 Sex: “Female”.

3.1.5. Case definition(s)

Case definitions for exposure and/or outcome should be produced.

Frequently only case definitions for outcome are needed and tend to be a MedDRA term, a list of MedDRA terms, including standardised lists such as Standardised MedDRA queries and/or a case outcome description (e.g. fatal).

Detailing the process of creating a case definitions for outcome is particularly relevant when:

- There is no single, agreed, definition of one concept or a corresponding MedDRA term or SMQ for that disorder (e.g. lack of efficacy);
- There is a need to discriminate between cases that only have reported the MedDRA term for the disorder from those that have the range of terms that match a certain diagnostic criteria (e.g. confirming whether individual cases match the RegiSCAR criteria for Drug reaction with eosinophilia and systemic symptoms) ;
- There is a likelihood of misclassification of the disorder;
- There is a need to define a precise indication for the use of a medicine (e.g. which terms coded in the indication field apply to a specific indication as per the SmPC, or when the analysis is to be restricted to cases of “normal conditions of use”).

Case definitions for exposure are usually limited to the substance(s) / medicinal product(s) of interest, but can include other criteria such as duration of exposure or exposure in patients with a specific medical history.

3.1.6. Analysis techniques

Analysis of EudraVigilance data may include qualitative and quantitative data analysis.

3.1.6.1. Qualitative data analysis

Some requests may require a clinical review of the individual case safety reports (ICSRs), which is usually the responsibility of the requestor. The purpose of the qualitative analysis is to facilitate the clinical interpretation of the cases rather than provide one. The exercise is distinct from the case review in signal detection as it is more about organising information rather than providing an opinion on the cumulative review.

This implies structuring the data to highlight the strength of evidence of each case preferably including an adequate line listing, and if needed, the ICSRs as annex.

Qualitative data analysis should prioritise primary sources of information, thus if an individual case review is warranted, and the line listing produced identifies literature references, the corresponding papers should be used as the primary source of information.

It is impractical to produce a narrative summary or a line listing for large numbers of ICSRs. In case of large ICSR numbers, the DAT should consider filtering by indication, seriousness, outcome, etc. The rationale for restricting cases should be described in the analysis report.

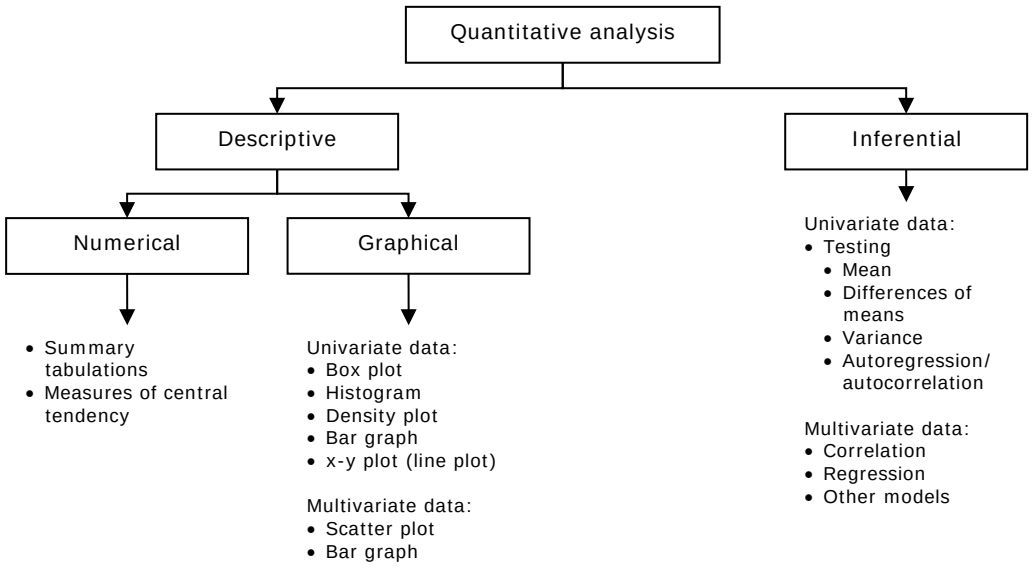
3.1.6.2. Quantitative data analysis

The quantitative analysis of a safety concern should start with descriptive statistics, i.e. summary tabulations, of the data.

Besides summary tabulations, a number of different types of quantitative analysis can be considered:

- Disproportionate analysis
- Time-to-onset analysis (including differences between subsamples and survival analysis)
- Time series analysis (including unexpected increase in frequency, changepoint analysis, outlier detection)
- Regression models (in particular multivariable logistic regression)
- Other models, including machine learning methods

A brief description of the types of analysis that can be performed, by type, is presented in the diagram below.



3.2. Quality control and validation

In a data analysis of EudraVigilance, there are three instances of quality control and validation: 1) Validation of the data extraction; 2) Validation of the code used in the analysis and 3) Validation of the results of the analysis.

3.2.1. Validation of the data extraction

To ensure that the raw data is correctly and exhaustively extracted, one DAT member should extract the raw data and another should validate the data extraction.

For each individual data frame extracted, i.e. a tabular data output with x rows and y columns such as a single table in an excel sheet, a different DAT member should compare the total count with the total count of his/her output and whether the unique identifiers are present. There is no need to run a full data extraction if they match.

For example, if DAT member A extracts a data frame with three columns, WWID, Age, Gender, and a thousand rows, DAT member B only needs to extract the WWID and compare if they are the same and that he/she has the same thousand rows.

Data extraction validation may sometimes be performed at the same time as additional data cleaning.

3.2.2. Validation of the code used (if appropriate)

If code is generated, it should be written in literate form, i.e. explaining what each line of code does. When code has been generated (e.g. R, Stata, SAS) this should be independently confirmed by a different team member, in particular when the results are inconsistent with other data obtained independently from the current analysis, e.g. a known predictor of the outcome or a known pattern in the data appears reversed in the analysis – such as higher frequency of breast cancer in males.

3.2.3. Validation of the results of the analysis

The peer reviewers do the validation of the results of the analysis.