# Summary:

The European Clinical Trials Database [EudraCT](https://eudract.ema.europa.eu/)  is run by the European Medicines Agency. All studies that are officially registered clinical trials have to enter the results of the final analysis into the website to be made public. There is a large amount of documentation online that will not be repeated here.

The detailed entering of safety information is an onerous task if it is to be done by hand. However we do have the facility to upload an XML file to automate this step. This tool seeks to enable the production of an XML file from a standard structure of Safety data listing that is recorded on an event and patient-level.

# Input Data

The tool will assume that the data provided is in the structure of a rectangular data file, with one row per adverse event, with the minimal set of variables

* Patient id
* Standardised description of the AE term, such as a MedDRA preferred term. Any conversion from free text is assumed to have been completed (for an example, term “Diarrhoea” and “Diarrhea”, or “Vomiting” and “vomiting” are read as different events).
* System Organ Class from MedDRA – a choice from 27 values (Appendix 1: eutctId)
* Serious: a logical value indicating if the event was serious
* Related: a logical value indicating if the event was related
* Fatal: a logical value indicating if the event was fatal
* Treatment group

This data records only events/patients that occurred, so if a patient had no AEs then they will not be present in the data.

So further statistics are needed:

* Numbers of patients exposed in each treatment group
* Numbers of deaths not reported in the AEs data , if there are any

# Outputs

## Step 1 Producing Summary Statistics

The EudraCTt database safety data Result requires the data to be summarised by counting the numbers of patients exposed and adverse events per term broken down by treatment group, and further sub-classified by seriousness, relatedness, fatality. Some further patients counts statistics are required on a treatment group level.

These statistics are normally calculated as part of standard clinical trial reports. Statisticians from different CTUs will most likely have scripts or libraries that produce these routinely, and will need to be tailored individually to fit the structure and variable names used within each trial. Hence we will only produce scripts for the main statistical packages (R, SAS, Stata) that produce rectangular data sets for examples and information.

## Step 2 Converting to XML and meeting the EudraCT Schema

The inputs to this step are to be three rectangular data sets

* group : one row per group providing group-level statistics
* non\_serious: reporting non-serious AEs only, one row per group-term
* serious: reporting SAEs only, one row per group-term.

### Group

The variable names will be exactly as below, case sensitive, to match the subsequent xml schema.

* title
* deathsResultingFromAdverseEvents
* subjectsAffectedBySeriousAdverseEvents
* subjectsAffectedByNonSeriousAdverseEvents
* subjectsExposed
* deathsAllCauses

title is the name of the treatment group; the meaning is clear of other variables, which are all integer counts of patients.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| title | deathsResultingFromAdverseEvents | subjectsAffectedBySeriousAdverseEvents | subjectsAffectedByNonSeriousAdverseEvents | subjectsExposed | deathsAllCauses |
| Arm1 |  |  |  |  |  |
| Arm2 |  |  |  |  |  |
| … |  |  |  |  |  |

### Non\_serious

This data set is rectangular and includes all possible combinations of terms (system organ class), and groups, even if in a treatment group there are no occurrences of a term. Note that a term is not required to be strictly nested within a unique system organ class, so the same term can be used with two distinct System Organ Classes, and they must be considered as distinct terms. However only combinations of term-system-organ-class that are observed need to be provided, rather than a full set of all possible combinations.

The non-serious data set is restricted to non-serious events

The minimal set of variables is:

* group
* subjectsAffected
* occurrences
* term
* eutctId

‘term’ is the description of the AE, such as the MedDRA

preferred term. ‘eutctId’ is the numerical coding for the system organ class as required by the Eudract Results schema. See Appendix 1.

If any filtering by event rate is required then this is assumed to have been applied.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| group | subjectsAffected | occurrences | term | eutctId |
| Arm1 | 5 | 8 | Nausea | 100000004853 or Gastrointestinal disorders |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

### Serious

This is an extension to non-serious but only reports on serious events. The same set of variables as for non-serious are required plus extra variables below:

* occurrencesCausallyRelatedToTreatment
* deaths
* deathsCausallyRelatedToTreatment

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group | subjectsAffected | occurrences | term | eutctId | occurrencesCausallyRelatedToTreatment | deaths | deathsCausallyRelatedToTreatment |
| Arm1 | 3 | 4 | Sepsis | Infections and infestations | 3 | 2 | 1 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

## Intermediate Basic XML data

Each of these three summary data sets will be saved in a basic xml format, in a generic grid structure:

<name\_of\_data>

<row>

<var\_name1>value1</var\_name1>

<var\_name2> value2</var\_name2>

…

</row>

<row>

…

</row>

…

</name\_of\_data>

An example is below

<group>

<row>

<title>Control</title>

<deathsResultingFromAdverseEvents>9</deathsResultingFromAdverseEvents>

<subjectsAffectedBySeriousAdverseEvents>15</subjectsAffectedBySeriousAdverseEvents>

<subjectsAffectedByNonSeriousAdverseEvents>15</subjectsAffectedByNonSeriousAdverseEvents>

<subjectsExposed>99</subjectsExposed>

<deathsAllCauses>12</deathsAllCauses>

</row>

<row>

<title>Experimental</title>

<deathsResultingFromAdverseEvents>22</deathsResultingFromAdverseEvents>

<subjectsAffectedBySeriousAdverseEvents>33</subjectsAffectedBySeriousAdverseEvents>

<subjectsAffectedByNonSeriousAdverseEvents>24</subjectsAffectedByNonSeriousAdverseEvents>

<subjectsExposed>101</subjectsExposed>

<deathsAllCauses>27</deathsAllCauses>

</row>

</group>

The three input data sets (group, serious, non\_serious) will be concatentated within one xml file within the root element <safety\_data/>

A validated script or function will be provided in R to save internal data into this format. Further tools within SAS and Stata may subsequently be provided and validated.

## Final conversion

An XSLT (<https://en.wikipedia.org/wiki/XSLT> , <https://www.w3.org/TR/xslt-10/> ) script will be provided to convert the intermediate basic xml document to a form compliant with the Eudract schema (<https://eudract.ema.europa.eu/docs/technical/schemas/clinicaltrial/results/adverseEvents.xsd> ). The XSLT script will be validated.

XSLT is a programming language specifically suited to converting the content of an xml data into whatever format and schema is desired. XSLT is a declarative language, which describes what the output is to be (rather than how to get there), making it very easy to amend and test. XSLT is neutral to the statistical package used in step 1.

The XSLT script and the input xml document will require an XSLT processor to perform the conversion. The processor will need to be provided by the end-user separately, and the validation of the processor (as distinct from the script itself) is the responsibility of the end-user

Numerous different XSLT processors are freely available (<https://en.wikipedia.org/wiki/XSLT#Processor_implementations> ), including ones supported by Microsoft, Java, Javascript, MacOS and a variety of linux distributions. Web browsers are able to apply the conversion, and there are XSLT processing add-on tools provided within statistical packages (R <https://cran.r-project.org/web/packages/xslt/index.html> , SAS PROC XSL). Version 1 standard will be used to ensure maximum compatibility.

# Limitations

In theory the following can be set at a default level for the trial, and altered from the default for specific AEs

* description
* assessmentMethod
* dictionary
  + name
  + version
* timeframe
* nonSeriousEventFrequencyThreshold

However, for the tool, default values will be applied with no ability to change within the tool. Rather any amendments needed are to be made by hand within the webpage after the upload.

# Validation

The validation will consist of a set of input data sets, and a corresponding set of output data sets, with an accompanying test document to explain how the outputs are judged as valid. Thus, the end-user will be able to perform User Acceptance Testing by following the steps in the test document, using the input data, and then compare the output obtained locally to those shipped with the tool. An R script will be provide to attempt to automate these steps as much as possible.

The components to be validated are

* Scripts in statistical packages to save the input data sets into the required basic xml format
* XSLT script to convert the basic xml file into the Eudract Results schema.

# Documentation

A user manual will be produced that attempts to guide a new user who has a working knowledge of a statistical package. In addition, there will be provided

* example input data sets
* scripts in statistical package languages for Step 1
* example intermediate data sets
* final outputs in xml and the pdf document produced by Eudract
* validation files (see Validation)

# Delivery

The documents, scripts and data described above will be made publically available within an online github repository (<https://github.com/> ). This provides a ready-made portal that provides strong version control and allows a community of users and developers to:

* download the tool
* comment upon or improve the tool, in a recorded and structured manner

# Appendix 1: eutctId

The coding of the system organ class is provided below

|  |  |  |
| --- | --- | --- |
| soc\_term | eutctId | meddra |
| Blood and lymphatic system disorders | 100000004851 | 10005329 |
| Cardiac disorders | 100000004849 | 10007541 |
| Congenital, familial and genetic disorders | 100000004850 | 10010331 |
| Ear and labyrinth disorders | 100000004854 | 10013993 |
| Endocrine disorders | 100000004860 | 10014698 |
| Eye disorders | 100000004853 | 10015919 |
| Gastrointestinal disorders | 100000004856 | 10017947 |
| General disorders and administration site conditions | 100000004867 | 10018065 |
| Hepatobiliary disorders | 100000004871 | 10019805 |
| Immune system disorders | 100000004870 | 10021428 |
| Infections and infestations | 100000004862 | 10021881 |
| Injury, poisoning and procedural complications | 100000004863 | 10022117 |
| Investigations | 100000004848 | 10022891 |
| Metabolism and nutrition disorders | 100000004861 | 10027433 |
| Musculoskeletal and connective tissue disorders | 100000004859 | 10028395 |
| Neoplasms benign, malignant and unspecified (incl cysts and polyps) | 100000004864 | 10029104 |
| Nervous system disorders | 100000004852 | 10029205 |
| Pregnancy, puerperium and perinatal conditions | 100000004868 | 10036585 |
| Product issues | 100000167503 | 10077536 |
| Psychiatric disorders | 100000004873 | 10037175 |
| Renal and urinary disorders | 100000004857 | 10038359 |
| Reproductive system and breast disorders | 100000004872 | 10038604 |
| Respiratory, thoracic and mediastinal disorders | 100000004855 | 10038738 |
| Skin and subcutaneous tissue disorders | 100000004858 | 10040785 |
| Social circumstances | 100000004869 | 10041244 |
| Surgical and medical procedures | 100000004865 | 10042613 |
| Vascular disorders | 100000004866 | 10047065 |