|  |  |
| --- | --- |
|  | **SOP Title:** Data Coding and Medical Coding |
| **Study title**: *Give study title to which this SOP applies* |

# Scope and application

|  |
| --- |
| Coding allows the reduction of large information into a form that can be more easily handled during data processing, reporting and analysis.  In clinical research there are two types of coding.  A first broad type is where a code (e.g. number) is used instead of writing the larger corresponding text. This type of coding is used during the CRF design and programming. This process will be referred to as ‘Data Coding’  A second type of coding is applied to medical terms and is called ‘Medical Coding’.  Medical coding is the process of coding medical terms such as Adverse events, Medical history and medication to a standardized format. Which terms should be coded depends on what medical data will be analyzed. This should be described in the protocol. This SOP will focus on the use of MedDRA for Medical coding, but other dictionaries also exist.  The aim of this procedure is to define coding and to provide instructions on the different aspects of coding. |

# Responsibilities

|  |  |
| --- | --- |
| **Function** | **Activities** |
| Data Manager | * Ensure that questions on data collection tools are correctly designed or programmed to perform coding * Coordinate the (medical) coding processes (collection, discrepancy management, coding maintenance) * Review that study data are properly coded * Extract the coded results from the database and provide the coded results AND coding dictionary to the statistician along with any clarifications where applicable. |
| System developer | * Ensure (if applicable) that the eCRF is correctly programmed to perform data and medical coding |
| Data Entry staff | * Respond to medical coding related DCRs |
| Data reviewer | * Work together with the medical coder to review the study data |
| Study physician | * If needed, give assistance or clarification on medical terms during the coding process. |
| Statistician | * Review that study data is properly coded * Extract study data and perform automatic matching of study data to the medical dictionary (If not possible in the CDMS) |
| Medical Coder | * Identify and agree on dictionaries to be used in the study * Reviewing the data to ensure it is ready for coding * Perform the automatic coding (if possible, in the CDMS) * Perform the manual coding |

# Definitions

**(e)CRF**: (electronic)Case Report Form

**CDMS:** Clinical Data Management Software

**DCR:** Data Clarification Request

**Medical Coding:** is the process of coding medical terms such as Adverse events and Medical history to a standardized format.

**MedDRA**: Medical Dictionary for Regulatory Activities. MedDRA is a medical coding dictionary developed and maintained by the Maintenance and Support Services Organization (MSSO). Supported by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)

# Procedures

**4.1. Data Coding**

* During CRF and database design, certain questions will be assigned with codes to categorize and standardize free text responses. These codes might be in number format (e.g. 1=Yes; 0= No; 999=Not known…) or text format (e.g. NA= Not Applicable; NK= Not Known; ND= Not Done…). In the database, codes should preferably be grouped in one table.
* Instructions on how to code data should best be guided via short instructions on the CRF or database or mentioned in data collection or data entry guidelines.

**4.2. Medical Coding  
4.2.1. The medical coder**

A medical coder should have a medical or life sciences background and ideally also at least some data management knowledge. They should also have a sound knowledge of the structure of MedDRA (or other medical dictionary used) and how to apply it for medical coding.

To avoid bias the medical coder should not be part of the data entry process and should not interact with study participants. They can however be involved in other aspects of the study (e.g. data management).

**4.2.2. Preparing the eCRF**

This step depends on the clinical data management software used for the study. If a software is used that can perform part of the medical coding it is important to design and program the eCRF accordingly.

If medical coding is done outside of the CRF no additional steps are required.

**4.2.3. Preparing the data**

The to be coded data should be thoroughly reviewed before coding. Medical terms should be checked for unclear terminology and abbreviations and queried where needed. It is also important that one medical term is listed per adverse event. For example, the term ‘nausea and vomiting’ should be split into two adverse events, as they cannot be coded together.

**4.2.4. Extracting the data**  
If coding is not done in the CDMS itself the data should be extracted by the statistician, using strict tracking and version control.

**4.2.5. Medical coding**

The medical terms will first be matched to the medical dictionary to find exact matches. This can be done in the CDMS itself (if possible) or by the statistician. All terms that do not have an exact match will be coded manually by the medical coder.

To ensure that the manual coding is done consistently a coding conventions document should be followed. This document describes how to prioritize in certain situations (e.g. a medical term where both an infection site and an infectious agent are mentioned eg. *S. aureus* infection on arm ).   
In case MedDRA is used, a coding conventions document, provided by MedDRA with each release, can be found at https://www.meddra.org/.  
If the study has very specific adverse events it is good practice to write a study specific coding conventions document to include study specific situations.

Where needed the medical coder can ask a study doctor or investigator for a second opinion or clarification of medical terms or medications.

The type of dictionary and the version and/or language should be tracked throughout the study. Either one version is used for the whole duration of the study, or data is recoded with each release. This should be documented in the Data Management Plan.

**4.2.6. Timelines**

Medical coding should be done as required throughout the study and should be completed before database lock.



# Attachments

|  |  |
| --- | --- |
| **Attachments** | |
| **Number** | **Title** |
| NA | NA |

1. **Document History and References**

|  |  |  |  |
| --- | --- | --- | --- |
| **Revision** | | | |
| **Version number** | **Author** | **Date** | **Description/reason for modification** |
| 1.0 | Hanne Landuyt | 04/10/2019 | Initial version  Review by Fatoumatta Cole and Yusupha Njie  Approval by Bai Lamin Dondeh and Harry van Loen |
| 2.0 | Hanne Landuyt | 20/06/2022 | Review to ensure that the SOP is appropriate within ALERRT and with current clinical research best practices. |

1. **Approval**

|  |  |  |
| --- | --- | --- |
| **Name and function** | **Date (dd/mm/yyyy)** | **Signature** |
| ***Author*** | | |
| *Indicate who wrote the SOP* |  |  |
| ***Review*** | | |
| *Indicate WP team members who reviewed (if applicable)* | *Date of review* |  |
| ***Approval*** | | |
| *Indicate WP Lead/Co-lead(s) who approved* | *Date of approval* |  |