|  |  |
| --- | --- |
|  | **SOP Title:** Data Validation & Review |
| **Study title**: *Give study title to which this SOP applies* |

# Scope and application

|  |
| --- |
| This SOP defines all the key aspects involved in data validation and reviewing of collected data in a clinical research project. While the objective of this process is not to specify a working method, it attempts to create a framework within which quality assurance is attained throughout the project.  The process of reviewing and validation is carried out periodically and as often as possible to maintain data quality throughout the study. Data validation can be carried out at different stages.   * During data entry: manual or automatic checks that are developed and embedded as part of the database e.g. edit checks. These may vary depending on the system used and the specifications drafted by the study staff. * Once data has been entered: systematic post entry checks need to be executed either automatically through an inbuilt module of the Electronic Data Capture (EDC) system being used e.g. REDCap’s Data Quality Module or OpenClinica’s Rule Designer or manually e.g. using SQL queries, SAS, R, etc. These must be defined and programmed before study start and executed at agreed specified intervals. * Ongoing monitoring: either members of the research team or independent monitors perform validation as part of the ongoing monitoring of the study. Such validation is done via Source Data Verification (SDV). SDV involves checking the data entered into the CRFs against that in the original source records e.g. patient’s hospital files for accuracy. * Quality Control (QC) and Quality Assurance (QA): QC calls for reviewing of specific aspects of the study data before moving on to the next step and QA tasks usually take the form of data management or database audits that are planned and then executed at specific time points.   This SOP applies to all clinical research projects where data is entered from paper forms into a database or directly into a database using an Electronic Data Capture (EDC) process. It also applies where mobile data collection equipment is used to capture data on/offline and imported into a database. |

# Responsibilities

|  |  |
| --- | --- |
| **Function** | **Activities** |
| Data Manager | * Defining and managing the process for data cleaning * Contribute to the development of the data validation plan * Review and approve the data validation plan |
| Project Lead (or delegate) | * Lead the development of the data validation plan * Review and approve the data validation plan |
| System Developer | * Program all items specified in the data validation plan into the data management system where possible |

# Definitions

**CRF**: Case Report Form

**Data Entry Checks**: All system in-built checks that are activated during data entry (automatic edit checks).

**DRL**: Data Review Log

**DVP**: Data Validation Plan

**ECS**: Edit Check Specification

**Post-entry checks**: All programmed checks that are applied to the data after data entry.

**SDV**: Source Data Verification

# Procedures

* 1. **Planning**
* Review the CRF design, study forms and study protocols and draft the Data Validation Plan (DVP). This is done by using the DVP template (see Attachment 01).
* Decide on the contents of the validation plan by:
  + Deciding what needs checking/validating
  + Working out when the check should be used
  + Specifying how to resolve inconsistencies and issues identified by the checks
* For data entry checks, draft an Edit Check Specification (ECS) to provide full details of the data entry checks that need to be set up and how each check will be tested. The ECS must be submitted to the database developer prior to or at the early stages of development of the database.
* For post-entry checks, develop a list of computer checks to identify missing data, perform context checks, logic checks, duplicate entries, etc.
* For SDV, agree on a percentage of records to carry out SDV on.
* Send the final draft of the DVP for review and approval.
* Ensure the approved DVP is referenced by the Data Management Plan (DMP).
* Create a Data Validation Checklist (see Attachment 02) based on the DVP.
  1. **Development and Testing of Post-Entry Checks**

Some systems may not have the ability for all checks to be embedded within the data management system. Therefore, alternative platforms external to the system may be required for e.g. SPSS, SAS, MS SQL Server, MS Access, R, etc.

* Using the DVP develop a query for each check using the platform of choice that is most suitable.
* Run a test on each check using the Data Validation Test Plan template (see Attachment 03). The process involves using test data to test for violation, fixing any violations found and re-running the test until all checks returns non-violation results.
  1. **Data Review Process**
* Carry out the data review periodically or as frequently as needed during the data collection phase. This is based on agreement with the study team ideally after each batch of data is collected or at specific agreed intervals.
* Carry out each check on the data as described on the DVP and record errors on the Data Review Log (DRL) (see Attachment 02).
* For each recorded error or clarification needed, a data query will be raised for investigation and correction by the study team. Details of this procedure are outlined in SOP-WP3-21-Query Management.
* At any time during or at the end of the study and before database lock, additional checks may be added and the DVP updated accordingly.
  1. **Final Data Review Process**
* At the end of the data collection process, ensure that a final data review process is carried out with no/zero errors on the outcome before the database can be locked, except issues noted and agreed with the study team.
* Using the DRL, perform a final run of all checks on the DVP.
* Complete a final DRL with the outcome section indicating zero count on errors. The only exception are insoluble queries that have been agreed as insoluble by the study team. Provide comments explaining why the errors are deemed insoluble.
* Sign and date the final DRL and commence preparations for database lock.
* Following a database unlock with changes, a final run of the checks will be carried out where necessary.

1. **Attachments**

|  |  |
| --- | --- |
| **Attachments** | |
| **Number** | **Title** |
| 01 | Data Validation Plan (DVP) |
| 02 | Data Review Log (DRL) |
| 03 | Data Validation Test Plan |

1. **Document History and References**

|  |  |  |  |
| --- | --- | --- | --- |
| **Revision** | | | |
| **Version number** | **Author** | **Date** | **Description/reason for modification** |
| 1.0 | Yusupha Njie Fatoumatta Cole | 04/10/2019 | Initial version  Review by Hanne Landuyt  Approval by Bai Lamin Dondeh and Harry Van Loen |
| 2.0 | Fatoumatta Cole | 23/06/2022 | 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* |  |