Standard Operating Procedure

Clinical Data Management

Identification No: DAT01

Version: 2

	Name	Role	Date	Signature
Written by	A. Carobbio	FROM Biostatistician		
Reviewed by	A. Masciulli	FROM Project Manager		
Reviewed by	T. Bordoni	FROM Quality Assurance Manager		
Reviewed by	S. Brusorio	FROM Clinical Operations Coordinator		
Approved by	E. Sfreddo	FROM Operational Director		
Approved by	A. Gavazzi	FROM Res. Team Coord. / Phase I Medical Director		
Approved by	F. Pezzoli	ASST-PG23 Chief Medical Officer		

TABLE OF CONTENT

1. ABBREVIATION

ASST-PG23: Azienda Socio-Sanitaria Territoriale - Papa Giovanni XXIII

CDMS:	Clinical Data Management System
CRA:	Clinical Research Associate
DB:	Database
DM:	Data Manager
DVP:	Data Validation Plan
eCRF:	electronic Case Report Form
FROM:	Fondazione per la Ricerca Ospedale di Bergamo
ISF:	Investigator Site File
OD:	Operational Director
PI:	Principal Investigator
PM:	Project Manager
SOP:	Standard Operating Procedure
TMF:	Trial Master File

2. SCOPE

This SOP defines the data management activities (e.g. CRF design, database development, clinical data review and handling, database lock) to be performed in relation to FROM sponsored or supported trials, and ASST-PG23 sponsored trials.

3. FIELD OF APPLICATION

This SOP applies to FROM and/or System Provider according to the agreement between the parties. This document should be used as guideline by the PM/DM/Statistician to better define the activities to be delegated to a System Provider, if any, and to monitor the data management activities performed.

4. RESPONSIBILITIES

Project Manager

- Coordinates the adequate management of clinical data of trials sponsored or supported by FROM.

- Supports test demo version of the eCRF.

- Provides the list of the FROM people who should be enabled to have access to the database.

- Assesses impact of protocol amendments on data management with the collaboration with the PI.

- Supports the draft sample CRF (paper) according to the protocol with the Principal Investigator and the Statistician.

Data Manager

- Performs and ensures data management activities as described in this SOP.

- Drafts data management documents (e.g. Data Management Plan, Data Management Report, Data Validation Plan) and ensures their approval.

- Supports Statistician to design, sets-up, test, validate the demo version of the eCRF and locks the trial database.

- Supports Statistician to prepare the eCRF guideline.

Statistician

- Reviews the draft "sample CRF" (paper) to ensure consistency with the statistical section of the protocol.

- Designs, sets-up, tests, validates the demo version of the eCRF and locks the trial database.

- Prepares eCRF guideline.

- Database access activation set-up.

- Interacts with the Service Provider for the data management.

Principal Investigator

- Approves data management documents.

- Drafts and approves the last version of CRF samples (hard copies design) with PM support, before their transposition into corresponding electronic version (eCRF).

- Tests demo version of the eCRF, approves and validate definitively it before put into production.

- Approves the data reported in the eCRF for each patient.

- Reviews data management documents issued by the DM (e.g. Data Management Plan, Data Management Report).

System Provider

- Release and maintains the trial database.

- Maintains documentation of the eCRF design process, key requirements, decision points and provides copy of these documents to the PM for filing in the TMF.

- Provides electronic copy of the concerned eCRF to trial sites and of the trial database to FROM at the end of the trial.

CRA
- Provides eCRF training to site staff.

5. PROCEDURE

The data management process covers the design and production of the eCRF, the design and construction of the database (DB), the processing of the data (entry, uploading cleaning and query management) and production of the final datasets ready for analysis.

FROM can delegate data management activities to a system provider. The contract in place between the parties includes details about this delegation.

This SOP section describes the main steps of the process of Clinical Data Management.

5.1 Data Management Plan and Data Management Report

As soon as the final draft Protocol of a trial is available, the data management activities can be started.

The documents describing all data management activities related to a clinical trial are the "Data Management Plan" and the "Data Management Report".

The "Data Management Plan" consists of an executive plan describing the systems, tools, criteria, procedures, responsibilities and timelines applicable to an individual trial; the "Data Management Report" includes a final summary of the activities carried-out during the trial, any specific issues emerged and relevant solutions, any unscheduled activities performed, and the final certification of data quality.

The drafting of the Data Management Plan and Report is performed by the DM and approved by the Principal Investigator.

The "Data Management Plan" should be amended and a new version should be issued whenever a significant change in data management procedures is required, (e.g., a protocol amendment which has a major impact on the eCRF and database structure) while minor changes are to be captured in the "Data Management Report" or in Notes to File.

5.2 System analysis and software selection

The activities described below will be performed by the DM.

Before starting any data management activity, the type of Clinical Data Management System (CDMS) to be used for the trial should be chosen. The CDMS consists of the database application and relevant integrated functions (e.g., query management). The selected CDMS should be validated.

The feasibility of its use should be verified at a given investigational site during the Pre-Trial Visit or by means of specific feasibility surveys, to ensure that all technical requirements are fulfilled and that appropriate site personnel are available for remote data-entry.

For each trial the appropriate clinical database location should be defined.

5.3 eCRF Design

The eCRF development and design process ensures that data elements required by the protocol are appropriately identified and electronic systems are in place for data storage.

The eCRF commonly is in a web-based system and are designed by the Statistician in a demo version and then deployed by the System Provider in the final web-based version of the trial database.

After protocol finalization and before starting database design a "sample CRF", based on the trial protocol, should be prepared. The "sample CRF" can be produced within the CDMS by the System Provider or as a separate document (e.g., using Microsoft Word or Excel) by the PI with PM support.

The draft "sample CRF" is reviewed by the Statistician to ensure consistency with the statistical section of the protocol.

The "sample CRF" is finally approved by the Principal Investigator by signing and dating the document **Modulo di approvazione della CRF elettronica (T.DAT01.01/2 – eCRF Approval Form)**.

The draft eCRF process should also include the following actions:

- identify the list of variables that need to be collected specifying the unit of measurement;
- design the eCRF layout deciding how the information should be included in the forms;
- define the *field type* (e.g. numerical, text, data, radio button, drop-down list, multiple choice checkbox);
- define the *field format* such as:
  - o date (e.g. mm/yy or dd/mm/yyyy);
  - o numerical (define number of digit and decimals);
  - o text (define the length);
- define the required formal check, if any;
- define the code list and code dictionary associated with the eCRF items.

Version control should be applied to the approved eCRF and all approved versions of eCRF used during the trial should be archived in the TMF.

5.3.1   Annotated Sample CRF and Data Dictionary

The following activities are delegated to the System Provider and reviewed by the Statistician.

 An Annotated Sample CRF is a "sample CRF" on which technical specifications are linked to each field to indicate the corresponding:

- database panel name and label;

- variable name and label;

- variable type (e.g., numeric, alphanumeric) and format;

- variable length;

- associated code lists (if any);

- source if the variable's content is derived from other variables.

The Annotated Sample CRF also includes in a separate section the list of codes and corresponding terms applied in the database.
The Annotated Sample CRF represents the executive plan for database programming.

The Annotated Sample CRF should be printed and archived in the TMF after database release as documentation of the database structure.

The database version number should be specified on the relevant Annotated Sample CRF and if the database requires to be amended during the course of the trial, the documentation of each database version must be produced and retained.

5.3.2   Procedure for eCRF Post-Production Changes

An eCRF post-production change is implemented if there was an error in the development of a form/data element or if an approved protocol amendment affects the eCRF design/data capture.

In case of protocol amendment the Statistician, PM, PI, DM and System Provider should assess whether the protocol change has an impact on data management, in particular on the eCRF, database structure and logical checks. If so the Service Provider ensures that the necessary actions are taken and properly documented.

5.4 Clinical Database Design and Training

5.4.1 Database Programming

The database used for the trial should be designed according to the sample CRF so that the data captured are complete, accurate, reliable and consistent.

5.4.2 Database testing

The following activities are delegated to the Statistician

eCRFs should be tested prior to trial start-up to evaluate the logic, wording and general construction of the eCRF as well as to verify that data can be properly entered and correctly captured.

This validation is performed in double-blind by two Statistician chosen among who did not designed the eCRF sample. These tests should be documented by screenshots of each component tested.

Once the eCRF is validated, a test has to be performed by the PM in charge to certify that the demo version is identical to the sample CRF and suitable for the trial.

At least one dummy eCRF is completed by the PM in all fields of the sample CRF to verify that expected data can be entered in the database exactly as they appear on the sample CRF (e.g. that numeric values of a certain expected length and decimals number can be entered in a given field).
A paper dummy CRF is used by the DM for testing the data entry in the database. If any data recorded on the dummy CRF cannot be entered or any other problems are encountered during data entry, the necessary changes are requested by the DM to and performed by the Statistician to correct the database structure.

An additional test has to be performed by the PI, in order to approve and validate the eCRF before its implementation as version ready to use.

The "validated eCRF" is finally approved by the Principal Investigator by signing and dating the document **Modulo di approvazione della CRF elettronica (T.DAT01.01/2 – eCRF Approval Form),** and it is also signed by OD, for acknowledgment.

The original paper sample CRF, completed and signed by the PM and PI, and the final print-out of the tested eCRF/database are archived in the TMF.

5.5 eCRF Completions Guidelines

A specific guideline for the e-CRF is prepared by the Statistician in collaboration with DM, and delivered to the trial site personnel and CRA(s). The guideline includes both technical instructions on the system use and trial-specific instructions on the e-CRF completion.

Any additional criteria for the CRF completion should be documented during the course of the trial in writing by the DM. Relevant information should be shared with all involved trial site personnel and CRA(s) to ensure consistency within the clinical database and correct interpretation of data.

5.6 Database Access and User Training

5.6.1 Database Users Identification

Database users are to be identified at System Provider, FROM and at all involved trial sites.

The trial team members are recorded on the **Trial Staff List (T.CLI05.01/2).**

Users at the trial sites are the PI and the personnel delegated by the PI to make entries and/or corrections in the eCRF. These people are identified by the PI and reported on the **Site Signatures and Delegation Log (T.CLI04.06/2**) to be archived in the ISF and a copy in the TMF.

The Statistician is responsible for assigning the database access privileges. The PM is responsible for providing the Statistician with the complete list of the authorized users with relevant privileges.

5.6.2 Database Access Privileges

The assignment of privileges is executed by the Statistician. The following database access privileges are usually foreseen in CDMS:

- data can be seen but not edited (i.e. "read only"). This privilege can potentially be granted to all personnel involved in the trial but for each trial site it is restricted to the data generated by that site;

- data can be seen and edited (i.e. "update", "read and write"). This privilege is granted to the PI and the personnel delegated by the PI to make entries/corrections in the e-CRF;

- electronic signature of the e-CRF (i.e. "approve"). This privilege is granted to the PI who takes responsibility for all data entered in the e-CRF at his/her site.

The Statistician is responsible for documenting in writing the specific privilege(s) assigned, any changes and the revocation for each user.

5.6.3 Database User Training

The database user training consists of a trial-specific training on the eCRF completion. It is provided to the CRA. In turn the CRA is responsible for training all people authorized to make entries in the database (i.e. PI and designees) and/or involved in the query process.

The training is documented on the **Site Staff Training Record (T.GEN03.02/2)** to be archived in the ISF and TMF.

5.6.4 Database Access Activation

Users are enabled by the Statistician to access the database only after appropriate training and the trial site personnel can be enabled only after the Site Initiation Visit.

5.7 Data Entry

During the trial data are entered directly into the database (e-CRF) at each trial site.

5.8 Randomization Procedures

In case of randomized trial design, the randomization procedure should be described in the trial protocol and further detailed in the Data Management Plan.

5.9 Data Validation Plan (DVP) and Automatic Checks Implementation

The Data Management Plan should state whether data review is performed with or without the aid of automatic checks.
If automatic checks have to be implemented, the DM prepares the Data Validation Plan (i.e. the list of control algorithms to be applied to the clinical database and relevant discrepancy messages) upon finalization of the sample CRF. Each control algorithm will be assigned a category as follows:

- Error: the result of this type of control is an error notification (e.g. missing or inconsistent data) or a warning for data that represents extreme observations (e.g. age = 99 years). The expected actions by the site personnel are a correction in the database in the first case and either a correction or a confirmation in the second case.

- Protocol Deviation: the result of this type of control is a protocol deviation notification (e.g. a mandatory examination was not performed as required by protocol or an entry criteria was not fulfilled). These notifications are to be reviewed by the CRA and submitted to the investigator and the expected action is the documentation of the reasons for deviation and relevant corrective measures taken/to be taken.

Errors provide a measure of the intrinsic quality of the data generated within a trial, while protocol deviations provide a measure of the quality of the protocol and of the trial conduct.
Each control algorithm is identified by a unique code within the DVP.

Automatic checks can be:

- Fully automatic checks: the electronically generated error/warning messages are addressed directly to the site personnel without prior review by the DM;

- Semi-automatic checks: the electronically generated error/warning messages are reviewed by the DM who decides whether it is appropriate to address them to the site personnel or not.

The final DVP is to be approved by the PM.

5.10 Data Cleaning

Data verification is a continuous process which can include three levels of control:

1) source data verification performed by the CRA to ensure consistency between the CRF and the corresponding source documents (this activity is out of the scope of the present SOP); this level of control is applicable to all clinical trials;

2) automatic (full or semi-automatic) checks performed by means of computerized programs aimed at ensuring completeness and logical consistency; this level of control is optional;

3) manual data review performed by the PM and DM. In case of medical issues, the PI is involved in the review. The level of control is based on appropriate risk evaluation (e.g. in case of very limited source data verification by the Monitor, the extent of manual data review should be greater).

Data cleaning continues until all discrepancies are resolved. If a discrepancy is accepted without changes to the database, a comment has to be reported by the PI to justify the reason for accepting the discrepancy.

5.10.1 Data Queries

Data changes resulting from queries are made by the site staff directly in the clinical database.

If the investigator's feed-back to a query is not a change to the database (e.g. a clarification or a confirmation of existing data), a comment is included in CDMS itself. If the CDMS does not support the management of queries, the investigator's feedback is reported on paper next to the relevant query and archived in the TMF.

5.11 Serious Adverse Events Re-conciliation

As a subset of clinical data concerning serious adverse events (SAEs) is recorded both on eCRFs and on SAE report forms, the PM or pharmacovigilance Service Provider must ensure that this information is reported in a consistent way. Even if aimed at collecting the same data, SAE report forms and CRFs often have a different structure and different completion criteria; therefore trial-specific re-conciliation rules and timelines are to be specified in the Data Management Plan.

5.12 Centralized Coding of Reported Terms

The list of variables to be coded centrally (i.e. the codes are not entered in the CRF by the site personnel) should be specified in the Data Management Plan.

Coding of adverse events according to the MedDRA dictionary is to be foreseen in all clinical trials with safety endpoints. The dictionary is included in the EDCS. The coding is performed firstly by the Investigator, then reviewed by the PM, who can require data queries, and finally checked by the pharmacovigilance Service Provider if appropriate (according to specific agreement).

Coding review and approval should be documented at least before each interim analysis (e.g. DSUR, presentation at scientific congress) and before database lock.

5.13 Handling of data from other sources

If the evaluation of the trial endpoints involves data that are not to be collected by the eCRF, the procedures for handling of this data are to be detailed in the Data Management Plan.

If the use of central laboratories or other central assessments are foreseen in a trial and/or the trial protocol includes ancillary studies which are not part of the clinical assessments carried-out by PIs, the Data Management Plan should specify if this data will be included in the clinical database and – if so – the procedures and technical specifications to be applied.

Examples of this data could be:

a) Central laboratory data to be loaded into the clinical database.

b) Centralized evaluation of radiological images which results are entered in the clinical database by a centralized data-entry.

c) Pharmacokinetic/pharmacodynamic data not to be included in the clinical database but to be presented in the statistical analysis.

d) Assessment of patient evaluability for the trial endpoints to be determined centrally.

Re-conciliation procedures relevant to data present both in the clinical database and in any external data set (e.g. demographic data) should also be described in the Data Management Plan.

5.14 Interim Database Freezing

Any foreseen interim analyses/reports have to be specified in the Data Management Plan. Details about the scheduled time points and procedures for database freezing and electronic archiving of datasets used for the analysis must be provided.

Any unplanned interim analysis should be mentioned in the Data Management Report.

5.15 Database Lock

It is the responsibility of the DM to inform Statistician when the database is ready to be locked.

At that point all accesses for data-entry will be disabled and this is to be documented in writing.

Before approving the database lock, the following items are to be solved by the DM:

- any unresolved discrepancies, including those relevant to SAEs re-conciliation;

- significant protocol deviations;

- final list of coded items;

- any other issues potentially affecting data quality or interpretation.

If an unresolved discrepancy is confirmed to be acceptable, it should be documented in writing that no action is required, while for major discrepancies which are not considered acceptable, queries are to be generated and database lock cannot be approved until their resolution. In this latter case the site personnel responsible for queries resolution will be enabled to make entries in the database; this authorization must be documented in writing. After the required changes are made, all accesses - except that of the PI of each trial site - are definitively disabled.

The last step before database lock is the PI signature (i.e. electronic) to confirm and take responsibility for the data generated at his/her site.

Database lock can be authorized only upon finalization of the Data Management Report by the DM.

Database unlock is to be performed whenever significant errors or omissions are discovered after database lock. The database unlock request can be issued by the DM, PI or Statistician and is to be carefully evaluated by all concerned parties.

If the decision is to unlock the database, the database items to be changed must be specified in a written form, the relevant accesses to the database are granted, data are changed and the database is re-locked with the same procedures described above.

5.16 Case Report Form Filing

The eCRFs are signed by the PI electronically. At the end of the trial a copy of the site-specific eCRF, including the audit trail, are provided on compact disks or equivalent device by the Statistician to all PI(s).

5.17 CRF and Database Long-Term Archive

The long-term archive of the clinical database is under FROM's responsibility and is kept at least for the period requested by the applicable law.

6. REFERENCES

- Guideline for good clinical practice E6(R2) (CPMP/ICH/135/95).

- Decreto Legislativo 6 novembre 2007, n. 200 - Attuazione della direttiva 2005/28/CE recante principi e linee guida dettagliate per la buona pratica clinica relativa ai medicinali in fase di sperimentazione a uso umano, nonché requisiti per l'autorizzazione alla fabbricazione o importazione di tali medicinali".

- Legislative Decree no. 211 of 24 June 2003 - Transposition of Directive 2001/20/EC relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for clinical use.

- ICH Harmonized Tripartite Guideline, Guideline for General Considerations for Clinical Trials E8, 1997.

7. TEMPLATES

T.DAT01.01/2   Modulo di approvazione della CRF elettronica (eCRF Approval Form).

8. VERSION HISTORY

Version	Date	Reason for revision
1	30 Sept 2016	Starting document.
2	15 Dec 2019	Changes in DM/PM responsibilities. Statistician's responsibilities added. Minor wording correction.