DIRECT Consortium: Data Management/Data Quality Guidance

Version 2: 02nd April 2013

Prepared by [redacted]

Contents

1. Introduction

The DIRECT consortium is committed to ensuring the highest standards of current good scientific practice for all the work carried out by, or on behalf of the consortium.

This objective will be accomplished if all of the following conditions are met:

1. Experiments are properly planned and performed.
2. Staff are competent, by education or training to carry out the experiment.
3. Materials used are of adequate quality.
4. The equipment and devices used function properly.
5. Results are correctly recorded, analysed and reported promptly.
6. Procedures are validated to a level appropriate to their purpose.

This guidance addresses our conceptual framework for achieving aspects of point 5 (Results are correctly recorded, analysed and reported promptly) and point 6 (Procedures are validated to a level appropriate to their purpose) and for compliance with ICH-GCP, with particular reference to data management.

In clinical trials, many of the functions relating to data management are standardised and subject to written procedures. However in a consortium of this type, implementing this type of structure may lead to the following difficulties:

- Burdensome processes inappropriate for non-commercial and non-CTMIP studies
- Lack of harmony with local or national frameworks and methods of working

For this reason this *guidance* is offered to clinical sites performing prospective clinical studies for consideration. Alternative approaches to address data quality management can also be considered and we strongly encourage the clinical centres to raise their suggestions for discussion.

2. Data management plans

Individual studies will not be expected to have a Data Management Plan. These are usually implemented in clinical trials and determine all Data Management Procedures from database setup until database lock. Instead we will have this guidance applicable to all DIRECT studies. This guidance is a living document; it can be amended to fit the actual requirements of the studies as they arise. The guidance will cover the following aspects:

i. General description of database design, data entry and data handling
ii. Specification of handling of missing values and replacement strategies for erroneous data
iii. Specification of the validations performed both manually and electronically
iv. Specification of data query generation and handling
v. Determination of use of coding dictionaries and coded items
vi. Specification of database quality control procedures
vii. Specification of electronic data transfer

3. Case Report Form (CRF) design and electronic CRF (eCRF) specification

For prospective studies data collection will frequently be performed in a custom-built database comprised of data fields represented in an electronic CRF (eCRF).

Source data verification (SDV) via monitoring performed by a Clinical Research Associate is not necessarily expected. However, where possible it is encouraged. The requirement and procedure for this type of data quality check is determined locally.

Quality assurance (QA) monitoring of studies in the clinical trial sense is not expected.
It is however expected that the conduct of all studies will follow the principles of GCP. Further comments about site to site audit are presented later in this document.

To facilitate design of the CRF, a designated CRF designer should be nominated from the WP task. The designated CRF Designer will endeavour to:

- Receive the final study protocol from all the relevant clinical centres.
- Review the protocol or protocols and determine which CRF modules and fields within modules must be created to collect the data specified in the protocol.
- Create prototypes of unique modules in an appropriate program.
- Provide the CRF Modules to the appropriate staff for review.
- Implement changes proposed during the review in the modular version of the CRF.
- Repeat the review cycle until no further comments are raised
- Create a full version of the CRF or CRF modules and circulate these to the relevant clinical centres.
- Transfer the CRF specifications to [redacted] to enable eCRF build.

In case changes to a finalized CRF become necessary eg. As a result of amendment of the protocol, the designated CRF designer for the WP task will:

- Assess the impact of the CRF changes.
- Change the CRF modules affected or add new modules as necessary.
- Provide the revised version of the CRF for review to the appropriate staff.
- Implement changes proposed during the review in the full version of the CRF.
- Repeat the review cycle until no further comments are raised.
- Inform all project participants that a new CRF version is available.
  Transfer the CRF specifications to [redacted] to enable eCRF build.

4. eCRF development and validation

[redacted] within WP5 implements an electronic version for each paper CRF (eCRF). The eCRFs are implemented to resemble the paper CRFs as closely as possible. Tooltips are added to guide the user to enter data in the required format. Required fields are defined by the CRF designer, and are highlighted in yellow in the eCRF. The CRF designer can also specify a range of allowed values for a particular field.
[redacted] validates each implemented eCRF by entering both allowed and non-allowed values to the eCRF fields, and verifying that the values are correctly stored in the database, and that appropriate error messages are produced when non-allowed values are entered. Users are asked to test the developed eCRFs to make sure that the correct fields are captured, required fields are defined correctly, and that pre-defined ranges for fields are correctly implemented.

5. Database design

Database design functions are carried out within WP5. The database design team will endeavour to:

- Specify individual types, lengths, formats, and entry features for every variable field of a data table according to the CRF modules.
- Annotate the variable and table names as well as the variable formats used in the database system to a blank copy of the CRF.

- Ensure that the general data management conventions on database design are regarded during database definition. This includes the use of standard variable names, as well as the use of numeric variables instead of alphanumeric variables wherever possible. Where no standard variable names exist, variable names can be chosen freely. However, over several studies variable names and types must be consistent.
- Set up database tables according to the data specifications.
- Create data entry screens and Database interface on the appropriate defined variable tables. These screens should be as close as possible to the underlying CRF to facilitate data entry and minimize data entry errors.
- Prepare test data CRFs to test the database for correctness. These test data CRFs should consist of dummy data, not actual patient data, and contain a couple of validation errors for subsequent validation testing.
- Test the data entry screens and the variables of the database by typing the test data CRFs into the data entry screens. The test data CRFs are subsequently compared against database contents for correctness of the data entry. This type of validation is carried out in collaboration with staff from the clinical sites.
- Amend any errors or inconsistencies found during test data entry.
- Repeat the test data entry process until no further errors are found.
- Document the database structure by downloading the database into an appropriate system and printing the structure.
- Release the study database for production.

If an amendment to the finalized Database becomes necessary, the Database Design team will:

- Assess the impact of the changes on the already entered data.
- Propose a solution to enable database change while ensuring that data integrity and validity is maintained.
- Stop data entry into the database while it is undergoing changes.
- Implement the changes to the database tables and data entry forms.
- Test the changed database tables and Data Entry forms.
- Document the amended database.
- Release the database for production.
- Inform the project team that data entry can resume.

Validation procedures will be reviewed by the study team on an ongoing basis to enable identification of incorrect discrepancies and procedures that may be no longer required or/and additional check required.

6. Training for all staff involved in the handling of clinical study data

For each WP task, a WP task Training Representative is named, and for each Centre, a Centre Training Representative is selected. The duties of a Training Representative include but are not limited to the following:

- Participates in teleconferences and meetings where handling of clinical study data are discussed (e.g. how to fill in a CRF correctly)
- Coordinates training of Centre Training Representatives
- Oversees that clinical study data are correctly handled
- Functions as an expert for the WP task regarding handling of clinical study data
- Communicates any issues arising to the appropriate parties

Duties for a Centre Training Representative include but are not limited to the following:

- Seeks clinical study data training from the WP task Training Representative
- Trains all personnel handling clinical study data at the Centre
- Functions as expert for the Centre regarding handling of clinical study data
- Communicates any issues arising in the Centre to the Centre lead and the WP Training Representative

[redacted] will arrange database training to the WP Training Representatives and the Centre Training Representatives upon request.

It is the responsibility of the every member of staff involved in the handling of clinical study data to:

- Maintain the highest quality of work and comply with ICH-GCP standards.
- Carry out assigned tasks according to the available guidelines (Study Protocol, Analysis Plan)
- Keep an updated Training record and document any training received in the Training Record.
- Keep a recent and updated CV within the Training Record.

7. Completion, storage and archiving of the paper CRF at the study sites

Paper CRFs must be completed where possible. All data fields should be entered. Every feasible effort should be made to provide accurate and complete data. Logic and common sense should be used to interpret information supplied by patients and to scrutinise information from other sources. Consistency should be a goal. To support this it is strongly encouraged that one person should work on one CRF. If a subject drops out of the study before completion of the study, an end of study page should be completed to identify the reason for this. All relevant CRF modules or visits must be completed.

The study staff must ensure that all CRF pages are stored in folders in an organised fashion with each subject's CRF modules kept together. At the end of the study the paper CRFs must be stored with the study files (Trial master file and Investigator Site File) and appropriately archived following relevant local procedures.

8. Data entry from paper CRF to eCRF

Data entry is not performed independently by two different data processors into two independent databases. Generally, data entry is performed by a single person.

It is critical to verify that the subject identification (ID) numbers and initials on the CRF pages are consistent throughout the CRF. This ensures that ONLY data relevant to the subject ID is entered.
The eCRF modules are entered separately for each subject ID and access requires the use of a secondary data field (date such as DOB).

It is the responsibility of each clinical centre to ensure complete data entry of all relevant and available data in a timely fashion (ideally within 1 month of a patient visit) and it is suggested to maintain a log or tracking/project management tool to monitor progress of this.

Text is to be entered as is it seen in the CRF (according to the below data handling conventions which note exceptions). No additions or replacements are to be made at the time of data entry. Unclear numbers or text should not be guessed by staff performing data entry. If an error is identified at the data entry stage this must be discussed with the research team and a change made to the paper CRF (dated and initialled, old text scored through). At all times concordance of the paper CRF with the eCRF must be maintained. Ideally data entry of a CRF for a single patient should be completed solely by one person in order to prevent modules of the CRF being omitted from the data entry process. Prior to submission, entered data pages should be checked for completeness and accuracy.

The following conventions apply:

No.	Classification	Data Handling Convention	Applied by
1	Misspellings	Obvious spelling mistakes and errors in grammar can be corrected	Staff performing data entry
2	Time	24:00 h will be entered as 00:00 h	
3	Investigator comments	Administrative comments of investigators will not be entered in the database	
4	Missing Visits	A visit will not be entered if no data has been filled in due to premature discontinuation of the patient	
5	Dates and Numbers	Irrespective of the format of the date on the paper CRF, dates entered into the eCRF must be recorded using the specified data entry format. Numbers should fall within allowable ranges.	
6	ND,NA,NK (Not done, Not applicable, Not known)	These abbreviations cannot be entered into numeric fields. For dates, 'Year NK', 'Month NK' and 'Day NK' are provided as separate boxes.	
7	Ambiguous Data	Ambiguous Data (e.g. ticking of 2 fields in a questionnaire) will be clarified with research staff prior to entry	
8	Unreadable Data	Unreadable data will be clarified with research staff prior to entry	
9	Ambiguous or erroneous Dates or Dates not known	Handling of incomplete dates: • If the year is unknown, enter '2012' and check 'Year NK' box • If the month is unknown, enter '01' and check 'Month NK' box If the day is unknown, enter '01' and check 'Day NK' box	

It is the responsibility of the person performing data entry to ensure that data entry has been undertaken for all relevant CRF pages.

9. Availability of the DIRECT website for data entry/sample registration

Periodically, the website will be taken off-line for maintenance, improvement or other changes in functionality. Clinical centres will be notified by [redacted] in advance of any down time and will be notified when the system becomes available for use again.

10. Checks post data entry and pre-submission to the database (via the DIRECT website)

Each clinical site should develop a suitable system for randomly checking a proportion of the entered data. For example, a person with experience of the study may regularly check the entered data against the paper CRF and take responsibility for correcting eCRF data that has been incorrectly entered or interpreted.

A suggested data quality check post entry and pre-submission will comprise of checking of:

- All Consent forms (signature and date; version number)
- 100% of all data for 20% subjects (e.g. every fifth subject submitted)

This data quality check should take place before submission of the eCRF data, ideally.

11. Use of filenotes to document deviations

A difficulty that arises with datasets in terms of evaluating data completeness is obtaining the knowledge relating to when data fields are missing or are generally anomalous. Within the DIRECT consortium, we will use a filenote system to capture data omissions and deviations. In some cases data omissions may be the result of early discontinuation of the study subject. Filenotes are not required to explain early discontinuation as an End of Study CRF page is generally used to document this outcome.

Deviations from the protocol or study related SOPs at the clinical site should be documented, acknowledged by the PI and reported to and acknowledged by the study manager, in a timely manner. The study manager should prepare a filenote to describe the deviation, and explain why it happened. This can be then be used alongside the data to assess anomalies of the data and to help explain these anomalies. A template for filenotes is available on request.

12. Changes to the eCRF by the clinical site after data entry

It is acceptable for staff from the research site to change the eCRF after the page or module has been submitted to the database.

Each site should nominate an experienced member of staff with administration rights for the DIRECT database. This person can place a submitted eCRF back to an editing status. Additional advice on how submitted eCRFs can be edited, can be requested from [redacted]. The database only stores the most recent data to be submitted. Multiple edits to submitted pages are possible.

13. How to handle missing data or unknown data

There is a distinction to be made whether a value is truly missing or whether it was never required to be completed. If a “Not known” checkbox is provided in the CRF structure, this should be used to identify missing data. If a “Not known” checkbox is not present, the data field should be left blank.

Sections that have been intentionally left blank (drop out subjects) will be left blank in the database. On the paper CRF, blank pages (e.g. pages not used) should be crossed through, signed and dated by the person filling out the paper CRF.

Clinical centres will need to put in place a process for ensuring that all relevant data for a patient has been entered into the eCRF.

14. Completion of review: eCRFs completed, samples completed

Within the status functions of the website it is possible to review each centre’s progress in terms of whether sample registrations are present for subjects as well as started and completed eCRF entry.

Centres will be asked, near to the end of the study, to verify that the missing eCRF or sample registration modules are as expected. This will ensure that missing data is not due to oversight of data entry.

15. Local normal ranges and out of range laboratory tests

Local normal ranges are not obtained from study centres and out of range laboratory values are not identified. Given the number of clinical sites it is impractical to implement a system that records and identifies out of range local laboratory values.

16. Electronic Data Transfer of Laboratory Data

All laboratory tests are provided in a specified format and uploaded to the database by the laboratory personnel. The supplier is responsible for assuring that the data in the transmission is an exact copy of the source data prior to transmission. The transfer will include all data available for all subjects.

Data suppliers are expected to periodically QC data they have entered. Correction or amendment of laboratory data will be done at the source of the data (i.e. by the data generators) and amended data will be re-transferred.

Analytical and –omics labs are expected to have their own written quality management platform. Analytical centres will keep calibration and QC data indefinitely. Calibration (traceable to a reference method where applicable) and controls are expected to be used. A separate document will be issued by the DIRECT consortium to provide further guidance on expectations.

17. Data Quality inspection team

A team of DIRECT consortium members will be nominated to undertake periodic review of the entered data. Between 4-6 DIRECT consortium members will be required to join the Data Quality inspection team. The Data Quality Team members will work with the clinical centres to resolve any queries arising from the data submitted by the centre. Each Data Quality inspection team member needs to sign the Data Access SOP.

[redacted] will export all data out from the database in flat files on a quarterly basis, and place it on a Windows desktop inside the Analysis Server. The Data Quality inspection team will review the data and identify potential errors, inconsistencies and missing data. Clarifications are emailed back to the individual centres. If data needs to be corrected, the centre that submitted the data will be responsible for performing all corrections.

Members of the Data Quality Inspection Team need to fill in a Data Access Request Form. This form should be separate from any Data Access Request Forms they might fill in for data analysis purposes. Members of the Data Quality Inspection Team are not allowed to distribute or utilize the data they receive for data QC purposes for anything else. All other uses require a new Data Access Request Form to be filled in.

18. Generation and Resolution of Discrepancies

The output/results created following an Inspection will be returned to the clinical sites for changes to be made to the eCRF by the person with administration rights. Data clarification forms will not be used, instead simpler email based structure will be used. Each data clarification will document:

- Subject ID number
- Visit / CRF module / Data field
- The problem identified, as clearly as possible

- Where a clear data handling convention exists this should be explained and appropriate training offered

All data clarification requests will be issued with a unique number and are logged in a Tracking log before being sent. Sites will be required to respond to these email requests for data clarification within 1 month and will be asked to report back whether changes to the database were made or were rejected. If a clarification is rejected a reason must be supplied. Reception and update of the database (resolution) is recorded on the tracking log.

Self-evident data changes can be implemented by [redacted] as and when required.

19. Site to site Audit or Inspection

It is acceptable for lead centres to request to visit a participating clinical centre for the purposes of audit or inspection. Local governance procedures should be followed. Preferably, this type of visit should be carried out by staff with experience of the particular study and the process of audit/inspection (by training or experience). The lead centre should prepare a written report to summarise the findings of the visit and circulate this to all study centres in the WP task.

20. Release of Data to the Analysis Server prior to Database lock

Data can be released to the analysis server at any time. However, data released prior to database lock is subject to change and has not necessarily been quality control checked or subjected to data clarification.

21. Database lock

Database Lock is the procedure taking place after all Data Management activities have been performed. Data in the database will then be set to read-only status in order to avoid any further changes.

Database lock can only be authorised when:

- All CRF pages have been entered into the database.
- All external data has been loaded onto the database where applicable.
- All checks/ validation of study data are performed.
- All Data Clarifications have been resolved and entered or closed.
- All reconciliation processes have been completed.
- Decisions to analysis populations and data sets are agreed.

22. Medical Coding

No medical coding is currently planned for the DIRECT prospective studies.

23. Analysis Plan

The Analysis Plan describes all activities for analysis of the data after database lock. It should be written and finalised prior to database lock to ensure the integrity of analysis decisions.

The WP Lead will:

- Request a member of the WP task to prepare an Analysis Plan (AP) based on the clinical study protocol.

The nominated person will:

- Describe in detail in the AP the variables to be analysed or summarised.
- Specify the planned analysis of all analysis variables at a level of detail greater than is possible in the study protocol, in particular for secondary variables.

- Specify the statistical models used for testing and estimation: criteria for statistical significance relative to the hypotheses being tested; handling of centre effects; robustness of models used.
- Specify the analysis populations; unequivocal definition of criteria for selection of patients for analysis populations
- Define the handling of missing values, outliers, and other data problems such as the timing of visits.
- Determine which strategy and which degree of analysis programming validation will be used.
- Distribute the AP for review.
- Amend the AP according to the comments received.
- Finalise the AP with the support of the WP lead.