

Data collection and management

High-quality processes to accurately capture and process clinical trials data are essential for ensuring that study results and conclusions are free from error and can reliably guide future health care decisions. The steps of the data capture and management process are:

Case report form design ⇨ database set-up ⇨ data entry ⇨ data cleaning ⇨ query resolution ⇨ database lock

CRF design

Case report forms (CRFs) are used to record the information required by the trial protocol. An ideal CRF is clear and well-organised, provides direction, and facilitates good data collection and easy data entry. CRFs should be designed and developed in consultation with the full study team (principal investigators, statisticians, data managers, trial coordinators, etc.). Refer to the CRF Design resource sheet on the Outreach web site for more information on developing high-quality data capture instruments.

Database setup

To achieve maximum quality control, the trial database used to store the data captured by the CRF should ideally contain the following features.

- Navigation system (front-end menu). This should display all forms available for a patient
- Tracking system (for paper forms). The entering of data into the database should be tracked and the current status (logged, queried, finalised etc) of all forms displayed with relevant dates (for example, date queries raised)
- Data entry screen. This should look as close as possible to the CRF for ease of data entry and to minimise errors. Small studies may use simpler layouts. If you are using an eCRF, online checks should be performed at the time of the data entry (for example, checks for out-of-range data).
- Complete and accurate representation of all CRF data when paper CRFs are used with an electronic database.

- Audit trail. All CRF databases must have an audit trail which allows any original or updated entry to be available, attributed to a particular individual, and timestamped.

Data entry

Data can be entered into the database using single or double data entry methods. Single data entry involves entering the data once, ideally by qualified and trained staff. The accuracy of data entered in this way is not guaranteed. Even with extensive consistency checks at time of data entry, batch checking after data entry is recommended. With double data entry, data are entered twice, ideally by two independent people into separate databases. Discrepancies between the two sets of data are identified and reconciled with the hard-copy CRF. This eliminates almost all keystroke errors; however, errors of interpretation are not always detectable.

Data cleaning

Raw data generally contain errors or problems such as missing values, transcription errors, inconsistencies and keying errors that require the data to be cleaned before any analysis can be performed. Data cleaning identifies these problems via integrity checks and logic checks. Inconsistencies between data collected in different parts of the CRF should be carefully checked for. Conflicting data can arise in cases where one question has a series of subquestions, when the same data are requested on more than one CRF page, or when a response to one question (or series of questions) implies that other questions have not been answered consistently (for example, a new concomitant medication is recorded that implies an adverse event has occurred but no information on the adverse event itself has been recorded in the CRF). Missing data or illegible text are other examples of data problems. Missing data can be minimised by instructing study coordinators to carefully review CRFs to ensure all required fields are completed. The risk of obtaining illegible data can be reduced by designing CRF items that are answered via tick boxes rather than free text.

Data collection and management

Query resolution

When the answer to a query is provided, either the original entry was correct and therefore the original entry remains unchanged or a new item of data is provided and is entered into the database, resulting in a database update. However, the original entry also remains in the database as part of the audit trail.

Database lock

When the data collection and data cleaning are complete, the database is locked for final analysis. This can be performed by taking a copy of the database or, in some commercial data-entry packages, an option for locking the database is provided.

Contact the Outreach team (trials@ctc.usyd.edu.au) or through the website (www.outreachclinicaltrials.org.au) for further advice on this topic.