DATA MANAGEMENT IN EPIDEMIOLOGICAL RESEARCH

K.M.Prasanna Kumar

Data management involves data collection, entry, verification and data retrieval when required. The aim of data management is to turn information from the subject into a report, efficiently and without errors. Data management is the most important aspect of any epidemiological research. The investigator is responsible for the collection, quality, recording, maintenance and retrieval of the data arising from the epidemiological study. All steps involved in data management should be documented properly. In any later auditing of the study by an independent authority, the documentation will allow them to perform a step by step retrospective assessment of data quality and study performance.

Before starting any epidemiological projects a " standard operating (SOP)" Document should be prepared. The SOP provides detailed written instructions for procedures and a general framework for the implementation of all the function and activities of each procedure.

The principal investigator for the epidemiological research should arrange for a "starters meet" with the staff involved like superiors, co-investigators, colleagues, nursing staff, technicians, to have a meeting at the beginning of the study or all personnel involved in the study and explain to every one the importance and scope of the study. Ascertain whether they have understood the purpose of the study, the design and the importance of the data they are to collect. Each person should feel that he/she is important and his/her role in the project is significant. The success of any project depends on the human resource management and the commitment of the work force.

The investigator should identify the right people for the right job and inspire them with a common goal. He should meet all the persons involved in the research and individually fix the responsibility as to what is expected of them. He should explain to them the protocol and give sufficient time for them to understand the protocol. It is better to go through the protocol in detail to ensure consistent understanding and clarify any doubts.

Once the protocol is finalized, a sample of subjects should be selected and a trial run should be carried out in the field. This trial run will help the investigator and personnel involved in the epidemiological research to get a feel of the study.

The data for the trial cases should be entered in the prototype of the case record forms. The person entering the data must be trained to enter the data properly and understand the significance of all the information collected. If there are any clarifications by the staff in the data entry, it should be done at the beginning of the study to avoid future confusion. Once the trial fun is over, any problems with data collection should be investigated and suitable modifications should be made in the case record forms

Case Record Form (CRF) is a record of the data on each subject in a study as defined by the protocol. The investigator must make every effort to maintain confidentiality of information about the subjects, including the subjects identity when checking documentation.

Once the project starts, a regular weekly or fortnightly or monthly meeting of the staff is necessary. This meeting has to be held on a fixed day and time convenient to all the staff to have maximum participation.

The agenda of such a meeting should incorporate:

- Number of case recruited
- Any problems in recruitment?
- Case record form verification – each case record are to be verified and if any corrections made to be confirmed with the signature of the investigator.
- Procedures confirmation- each person involved in the project must explain how exactly he/she collected the data. What questions were asked and the interpretations of the same .
- Informed consent – if the study involved informed consent of the subjects, each consent form should be verified and made sure it is in proper order.
- Any problems in collecting the data must be investigated. Individuals to report any problem in collecting data, any other requirement or facility that is lacking in the field.
- Suggestions from the staff – opinions or suggestions from each member of the research team should be invited as to how the system can be improved. Any good suggestions should be implemented. There should be a continuous improvement program (CIP) in any work.

From Dept. of Endocrinology & Metabolism, MS Ramaiah Medical College, Bangalore 560054.

Data Collection & Entry: The data may be entered directly on to a computer or collected on the CRF and then fed in. Data entry should be performed continuously during the course of the study. It should be checked either by double entry i.e. computerized checking or by proof reading i.e. checked manually. If the data checking has to be done on computer, it can be combined with data entry in order to speed up feedback. To supplement the continuous checking of each individual's data during the study, descriptive statistics on each important variable in the database are useful in the detection of doubtful or unusual data.

Data Corrections: All corrections on the case record form and elsewhere in the hard copy data must be made in a way which doesn't obscure the original entry. The correct data must be inserted with the reason for the correction, dated and initialled by the investigator. If the data are altered during processing, the alterations must be documented and the system validated. For electronic data processing only authorized persons should be able to enter or modify data in the computer and there should be a record of changes and deletions.

Verification / Validation of the data: Data Quality assurance. The case record forms or computer data entries should be verified from time to time to ensure proper data entry during weekly or monthly meeting with research staff. The aim of data quality assurance is to minimize the effects of missing and inaccurate data. The data editing process includes defined procedures for confirmation and if necessary, correction of the data. The procedure followed for data editing should guarantee a rapid feedback to ensure that the process is efficient in bringing queries about validation to the attention of the investigator. Procedures should be designed and carried out to ensure that the data contained in the final clinical trial report (Final report) match original observations. These procedures may apply to raw data hard copy or electronic CRF, computer print outs and statistical analyses and tables.

Subject Identification List: This document is a list of the code numbers of the case record forms and the corresponding personnel identification numbers or addresses of the subjects participating in the study. This list should be maintaining by the investigator to make it possible to trace and identify the individual subjects in an epidemiological study. This list will help to carry out the medical auditing or inspections.

Laboratory Values with normal reference ranges should always be recorded on case record forms or attached to it. Values outside a clinically accepted reference range or values that differ significantly from the previous values must be evaluated and commented upon by the investigator. Data other than those requested by the protocol may appear on the case record form clearly marked as additional findings and the investigator should describe their significance.

Units of measurement must always be stated and transformation of units must always be indicated and documented.

Data Access: The investigator should retain the records and data from the study for safety for safety reasons and for auditing and inspection subsequent to study completion. It is better to store the data documents for at least 15 years all the original case record forms and data including lab reports, X-rays, microfilms, angiograms should be preserved in a water proof, fire proof cabinet. The key to this cabinet should be accessible to the investigator or any person authorized by the Investigator.

Access to data should be available only to authorized persons involved in the study. Confidentiality of the database is secured by appropriate standard operating procedures including passwords for all staff involved.

Satisfactory maintenance and backup procedures for the databases should be provided so that in case of hard disk problems in future the data is recoverable. If the data are entered directly into the computer there should be adequate safe guard to ensure validation including a signed and dated printout and backup records. Archived data may be held on microfilm or electronic record, provided that a back up exists and a hard copy can be obtained from it if required.

Transparency, authenticity and accuracy are very important in any epidemiological research. It is the responsibility of the investigator to ensure that the principles of good clinical practice are adhered to ensure the same.

References for Further Reading

1. GCP (Good clinical Practice) worldwide guidelines and regulations Australia, Japan, and Nordic, 1993, Eli Lilly and company, Book 1 & Book 2.

2. GCP (Good Clinical Practice) worldwide guidelines and regulations U.S. and E.C, January 1993, Eli Lilly and company, Book 1 & Book 2.