Guidelines for
DATA QUALITY ASSURANCE IN CLINICAL TRIALS AND OBSERVATIONAL STUedes

National Heart, Lung, and Blood Institute
National Institutes of Health

Revised: April 24, 2001

During the conduct of clinical trials and observational studies, two types of data errors can occur. One type is due to deliberate falsification; the second type results from human or measurement error (such as inaccurate data entries, inaccurate transfers of data, misinterpretation, and inherent limitations of the measurement instruments). Errors resulting from falsified data are always serious and must be dealt with accordingly; other errors may be serious or trivial. However, every clinical study must include procedures to avoid or minimize data errors. The focus of this document is on methods that are applicable to multicenter clinical trials, although many of the issues are just as relevant to single center studies.

Several changes have occurred over the years which make attention to data error more of an issue. In the past, the typical multicenter study had a number of well-funded clinics, selected at least partly on the basis of research experience, and a well-funded coordinating center. These features facilitated understanding and implementation of the protocol and the scientific method used by the clinic investigators. Also, with larger clinic staffs, several staff were around to catch both types of errors. Today, for clinical trials in particular, there is a greater
use of the so-called "large simple study" model. Here, almost any investigator with certain minimal qualifications enters participants, and payment is typically done on a per participant basis. As a consequence, there are many more investigators, there is less on-site monitoring, and each investigator may enroll fewer participants on average. Thus, the protocol may not be as well understood. Nevertheless, clinical trials and observational studies involve the collection of hundreds of thousands of data items. With care and appropriate (and sometimes costly) procedures, errors can be minimized, though not reduced to zero.

**Objective**

The objective of these Guidelines is to promote quality assurance by identifying various methods for reducing important errors such as phantom participants, fabrication of data, entry of participants with major ineligibility criteria, randomization mixups, and outcome measure mistakes. It is recognized that there are other kinds of errors which cannot be prevented or discerned, even if there were far greater resources and one could compare all study forms with the original records. In addition, not all data on the study forms come from medical records; many items are entered directly on paper or computer at the time of a participant visit to a clinic. Therefore, priorities must be set, and all appropriate efforts should be made to reduce the likelihood of important errors which compromise the essential integrity of the research data.

These guidelines are general in nature, as the specifics will vary depending on the nature of the study. They are divided according to the locus of responsibility.

**Data Center/Study Coordinating Center**
The Data Center/Study Coordinating Center should be an independent unit in the study organization. The Principal Investigator or the Steering Committee Chair of multicenter studies should not be in a position of authority over the data center. This autonomous arrangement may not be feasible in single center studies.

As a part of the peer review process, all data centers must demonstrate an understanding of the mechanisms of data quality assurance and indicate an intention to carry them out. Typically, data centers set up or coordinate training sessions which should address the following areas:

Extensive education and training of investigators and other clinic personnel, both at the start of a study and at intervals during the study, are essential to comply with the protocol. Training should include not only discussions of the study protocol and sessions on how to complete forms, enter data, and perform procedures, but presentations of basic concepts of research methods and bioethics.

The examination of data by clinical unit or site to identify possible outliers, both in participant enrollment and data quality, should be done routinely.

The examination of photocopies of the original forms for missing, outlying, and inconsistent data needs to be done regularly. With many studies using (computer) distributed data entry, some error checks will occur at the time of data entry. It must be recognized that there is no way of identifying some possibly legitimate but erroneous data, such as a translocation of numbers (e.g. blood pressure of 124/84 mm Hg rather than 142/84 mm Hg).
The comparison of collected data with photocopies of the original data sources, such as hospital or other medical records, laboratory records, and death certificates should be done selectively. This may be accomplished in several ways. For the more important items, a higher proportion, if not all, of the data should be confirmed by comparison with the original records. This can be done either through site visits or through requests for photocopies of sample medical records or other data which are sent by mail or fax to the data center. The size of the sample to be compared will depend on the nature and size of the study and will vary according to size of clinic and kind of form. If other monitoring techniques indicate a problem, or if there is an enhanced level of suspicion for whatever reason, the sample may well be 100 percent for a given clinic. Outcome forms should also be more closely examined than other forms, although this too will depend on the nature of the study (e.g., blinded or not). Collection of death certificates and hospital summaries for outcome events should be done and is essential for trials which have outcome classification committees. In some studies, sample collections of outcome validation records may suffice, but it is important to consider why all records should not be obtained.

If there are central laboratories, the Data Center needs to monitor quality control, including evaluation of missing or delayed data, outlying values, and temporal changes. It may institute procedures such as blinded resubmission of materials on a sample basis. The details obviously will be determined by the nature of the particular study and the analytical activities of the laboratory. Of course, each laboratory must have its own quality assurance procedures.

Internal quality control is an integral function of all Data Centers. Periodic checks of data flow, processing time, and data entry errors are important, as are reviews of analytic
procedures. Occasional site visits by the NHLBI project officer/program administrator, with or without external experts, are necessary.

**Clinical Unit**

Current NIH guidelines (http://grants.nih.gov/grants/guide/notice-files/not-OD-00-039.html) require that all clinical investigators supported by the NIH be certified as having had training in human subject protection.

In addition, the clinical unit Principal Investigator must exhibit a knowledge of clinical research approaches and an understanding of the goals of the protocol. He/she must clearly communicate these to clinic staff and assure staff compliance. As discussed above, education and training are essential. It must be emphasized that in clinical trials, an investigator who has a strong prior preference for a particular outcome should not participate in the trial.

It is the responsibility of the Principal Investigator to oversee staff activities regularly, ensuring high quality and timely conduct of procedures and data entry, and maintaining communication with the central units of the study. The Principal Investigator or designee should sign all completed forms, acknowledging responsibility for the data.

It should be mentioned that in most drug company sponsored trials, staff (generally nurses) are employed to travel to clinics to check every data form against the original medical records. This degree of validation has not been implemented in NHLBI studies, nor is it proposed in this document.

**Investigator Meetings and Committees**
In multicenter studies, regular investigator meetings allow for discussion of issues and education and training and are essential for promoting compliance with the study protocol and maintaining study integrity.

Multicenter studies generally have a committee structure which includes a steering committee with overall scientific management responsibility and specific committee(s) with quality control and protocol compliance responsibility. Thus, the typical study administrative structure may consist of a steering committee, executive committee, publications committee, and special protocol committees, i.e., special treatment committees. Regular reports (in clinical trials without outcome data by treatment group) are provided and serve to keep investigators attuned to the need for high quality valid data.

**Data and Safety Monitoring Boards (DSMBs) & Observational Study Monitoring Boards (OSMBs)**

The primary purposes of the DSMBs and OSMBs are to assure to the extent possible, participant safety and the scientific integrity of the study. Although the study investigators/Steering Committee have the primary responsibility for the integrity of the study, DSMBs and OSMBs carry out an important oversight role. The establishment and responsibilities of these Boards are described in the following documents located on the NHLBI home page.

[Establishing NHLBI DSMBs and OSMBs](http://www.nhlbi.nih.gov/funding/policies/dsmb_est.htm)
Responsibilities of DSMBs Appointed by Participating Institutions
http://www.nhlbi.nih.gov/funding/policies/dsmb_othr.htm

Responsibilities of DSMBs Appointed by the NHLBI
http://www.nhlbi.nih.gov/funding/policies/dsmb_inst.htm

Responsibilities of OSMBs Appointed by the NHLBI
http://www.nhlbi.nih.gov/funding/policies/osmb_inst.htm

NHLBI Project/Program Officer

The NHLBI project/program officer is responsible for the oversight of the conduct of the trial. In terms of data quality assurance, he or she obviously must delegate responsibility to other units such as the data center. However, in conjunction with the DSMB or OSMB, the NHLBI project/program officer will ensure that the data center appropriately carries out the quality assurance function.

Site visits to clinical units are necessary to assure data quality, but the number of centers visited and the frequency will depend on the nature of the study and the number of centers. There are many reasons for conducting a site visit. These include particularly poor participant recruitment as well as surprisingly good recruitment, and data quality problems noted as a result of data center monitoring or direct communication with the clinic. Site visits are essential if there are questions about data integrity.

As noted above, site visits to the Data Center are important. In addition to site visits, regular review of procedures and data tables should be done. For clinical trials, these tables may not
contain outcome data, which are confidential, but other sorts of data should be examined. The Data Center staff should bring all concerns to the attention of the project/program officer; the project/program officer should also monitor the Data Center, and remind its staff to do so as needed. The project/program officer must receive frequent tabulations of study progress such as participant enrollment.

Frequent communication by telephone, mail, e-mail, and fax is a necessary feature of all multicenter studies. A part of this communication should focus on clinic improvement in participant enrollment. The NHLBI project/program officer needs to balance the use of measures to improve enrollment versus inappropriate pressures which might lead to enrollment of ineligible participants.

With respect to large simple clinical trials, communication and quality control of the study through site visits are more difficult and thus require vigilant surveillance of data quality. The use of regional coordinators for studies with numerous investigators has proven to be useful, at least in reinforcing the protocol and improving participant enrollment. Double blinded studies can reduce bias, and they may be particularly useful in large, simple trials, especially if the primary response variable is anything other than all cause mortality.

For questions or comments regarding this document, please contact Dr. Lawrence Friedman, Special Assistant to the Director, NHLBI (friedmanl@nih.gov).