

# Sleep AHEAD Study

---

## Standard Operating Procedures For Data Management and Statistical Analyses December 3, 2001

### Table of Contents

<b>1.</b>	<b>Introduction.....</b>	<b>2</b>
<b>2.</b>	<b>Design and Management of Clinical Data Bases.....</b>	<b>3</b>
2.1	Purpose.....	3
2.2	Access Data Entry System Requirements.....	3
2.3	Handling of confidential identifying information.....	3
2.4	Data dictionary / codebook creation.....	3
2.5	CRF receipt log.....	4
<b>3.</b>	<b>Processing of Case Report Forms Requiring Manual Data Entry.....</b>	<b>5</b>
3.1	Review by Site Clinical Coordinator:.....	5
3.2	Initial Receipt of CRFs from Clinical Sites:.....	5
3.3	PDM Data entry and review:.....	5
3.4	Handling Data Questions:.....	5
3.4.1	Noting Data Questions:.....	5
3.4.2	Processing Data Questions:.....	5
3.5	Self-Evident Corrections.....	6
3.6	Double Data Entry by a study Data Clerk (DC).....	6
3.7	Double Data Entry Comparison.....	6
<b>4.</b>	<b>Receipt of Electronically Transmitted Data.....</b>	<b>7</b>
<b>5.</b>	<b>Statistical Screening and Statistical Data Management Functions.....</b>	<b>8</b>
5.1	Assessment of Data Quality.....	8
5.2	Assessment of Protocol Violators.....	8
5.3	Ongoing Study Management Support.....	8
<b>6.</b>	<b>Clinical Study Report Documentation.....</b>	<b>9</b>
6.1	SAS to Excel Updateable Tables.....	9
6.2	Proc Report Listings.....	9
6.3	SAS Listings.....	9
<b>7.</b>	<b>Backup procedures.....</b>	<b>10</b>
<b>8.</b>	<b>Forms Appendix.....</b>	<b>11</b>
8.1	CRF Data Query Form.....	11

## **1. Introduction**

A Standard operating procedure (SOP) is a document that specifies all the operational steps, acceptance criteria, personnel responsibilities, and materials required to accomplish a task.

SOP's are necessary in order to efficiently allow multiple staff members contribute to the progress of the study, to permit efficiencies in programming documentation and validation and to provide a framework for the ethical and scientifically sound conduct of human subject research. Written SOPs facilitate the highest quality study management. The procedures described in this document should be used by all Sleep Apnea in Look AHEAD (Action for Health in Diabetes) data management and statistical personnel in order to maintain optimally efficient operations and to comply with regulations and contracted responsibilities regarding standardization of operating procedures.

It is the intent of these SOP's to be consistent with Good Clinical Practice (GCP), a standard established by the International Conference on Harmonization (ICH) for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of clinical studies that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of study subjects are protected.

## **2. Design and Management of Clinical Data Bases**

### **2.1 Purpose**

The purpose of this section is to describe the system to be used to develop, monitor, manage, and assess the quality of designed and managed clinical databases to be used in this study. These procedures are designed to be consistent with Good Clinical Practice (GCP) and all applicable sections of the Code of Federal Regulations.

### **2.2 Access Data Entry System Requirements**

The following elements will be included in all Access databases designed for manual data entry.

- Automatic time and date that a new record was added
- Initial of data entry personal
- Password protection
- Form revision number field
- Pull down menus containing valid codes for fields as appropriate

A separate Access table for each data source will be created.

### **2.3 Handling of confidential identifying information**

Access data bases developed for use as input to statistical analysis databases will contain no personal identification information such as last name, telephone number, address, etc. A separate confidential information data base will include these fields and will only be maintained on removable computer media (either Zip drive or CD-RW). This removable media will be kept under lock and key and will only be accessible by the primary data manager and the Principal Investigator of the Data Coordinating Center. Should it become necessary, confidential information can be merged with temporary analysis datasets for specific purposes such as identifying patients with specific attributes. Such data sets will be constructed for a single run and then automatically erased. Hard copy files containing confidential information will also be kept in locked files cabinets accessible by only the primary data manager and the Principal Investigator of the Data Coordinating Center

### **2.4 Data dictionary / codebook creation**

The Primary Data Manager (PDM) will create a data dictionary / codebook. The data dictionary / codebook will contain:

- A blank copy of the case record form (CRF) with the Access variable names for every variable as well as the codes assigned to variable category levels written in.
- A coding guideline / coding convention document listing all guidelines and conventions developed at the start of the study or during the conduct of the study. Decision dates will be recorded.

## **2.5 CRF receipt log**

The primary data manager will create a CRF receipt log. This log will include the following elements. The log will be completed prior to data entry.

- Initials of PDM
- Study ID
- Date received
- Form name
- Form revision

### **3. Processing of Case Report Forms Requiring Manual Data Entry**

#### **3.1 Review by Site Clinical Coordinator:**

All clinical sites will have a designated Clinical Coordinator responsible for transmittal of CRF's to the Data Coordinating Center. Prior to transmittal, these CRFs will have been checked by for completeness and accuracy by the Clinical Coordinator.

#### **3.2 Initial Receipt of CRFs from Clinical Sites:**

The date of receipt will be documented on the CRF upon receipt by the Primary Data Manager and must not obscure any clinical data on the CRF.

The CRFs received by the PDM will be recorded on the CRF Receipt Log (defined above).

#### **3.3 PDM Data entry and review:**

During data entry by the PDM, CRFs will be reviewed thoroughly for the following:

- A. The CRFs must be complete, including identification of subjects by subject number and initials. Missing data are to be noted by the PDM.
- B. Data must be appropriate for the parameters requested. Data inconsistencies and protocol deviations are noted by the PDM.
- C. Data must be legible and capable of being photocopied (black ink preferred). Illegible data are noted by the PDM.

#### **3.4 Handling Data Questions:**

##### **3.4.1 Noting Data Questions:**

Upon review of the CRF, the PDM notes any data questions (i.e. items on the CRF that are missing, inconsistent, illegible, or need clarification) by completing a CRF Query Form (see Appendix A).

##### **3.4.2 Processing Data Questions:**

The PDM processes the data queries in the following manner:

**Sending CRF Query Forms:** The CRF Query Form is sent to the Site Clinical Coordinator by way of E-mail or fax. If urgent, the data query can be handled by a telephone call to the site Clinical Coordinator. The data response is documented on a CRF Query Form by the Clinical Monitor and sent back to the PDM.

**CRF Edit Method:** The site Clinical Coordinator responds to a data query by indicating the action to be taken by the PDM on the CRF Query Form. It is the responsibility of the Clinical Coordinator to resolve queries. CRFs requiring changes in information will be edited by the PDM. The incorrect response will be stricken with a single line through it and the corrected response recorded

next to it. The edit will be initialed and dated by the individual making the edit. A copy of the edited CRF pages along with the CRF Query Form initiating the required change will be faxed to the Clinical Coordinator. The Clinical Coordinator shall ensure that all edits are recorded on all other copies of the CRF. Data queries do not need to be resolved prior to data entry.

### **3.5 Self-Evident Corrections**

Self-evident corrections are those that the PDM may make to CRF data that are clearly wrong. The purpose of self-evident corrections is to avoid sending queries to the Clinical Coordinator for issues where the correct answer is obvious. A log of self-evident correction types will be compiled. This log will list the types of self-evident corrections made but not necessarily each and every occurrence. The incorrect response will be stricken with a single line through it and the corrected response recorded next to it. The edit will be initialed and dated by the individual making the edit. If a blank entry is added, the response will be recorded on the CRF and initialed and dated by the individual making the edit. An example of self-evident correction is adding information available on other forms that was left blank on a particular form.

### **3.6 Double Data Entry by a study Data Clerk (DC)**

After the PDM has completed data entry and recording of Data Queries, CRFs will be provided to a study data clerk (DC) for double data entry. Separate Access databases will be used in the double data entry. The Data Clerk will manually enter the data to the best of their ability but will not formally document questionable data fields.

### **3.7 Double Data Entry Comparison**

The study statistical programmer will develop programs that convert the primary and secondary Access databases into SAS datasets and perform the SAS Proc.Compare implementation of a field by field comparison. The program listing will be automatically saved with the run date defined as part of the filename in order to facilitate documentation of the comparison procedures.

#### **4. Receipt of Electronically Transmitted Data**

Data from the PSGRL will be transmitted to the Data Coordinating Center every other week. A SAS program will be written that extracts necessary variables. A similar procedure will be put in place for electronic data received from the parent study.

## **5. Statistical Screening and Statistical Data Management Functions**

Statistical screening and data management will be carried out as part of the statistical and statistical programming functions.

### **5.1 Assessment of Data Quality**

The statistical screening function in data quality assessment differs from that of the PDM. Statistical screening function does not verify that data in the clinical study database matches data on the CRF or if it is medically interpretable. Statistical screening involves reviewing data for aspects of data relevant to the statistical methodology that will be employed in the statistical analysis. As such, statistical screening includes analyses designed to uncover data “oddities” with special attention paid to primary efficacy and safety data. Exploratory data analyses of these data may uncover outliers and inconsistencies that do not meet acceptability criteria. Any identified problems will be forwarded to the PDM who will then confirm accuracy of data entry by examination of the CRF. If data entry is accurate, the PDM will confirm values through the Data Query process. Statistical screening listings will be designated as Report 0 or Report 0a, 0b, etc.

### **5.2 Assessment of Protocol Violators**

The second role of statistical screening involves identification of protocol violators. Identification is done by specially designed programming queries or through examination of patient listings. When protocol violators are identified, the data on which this determination is based is forwarded to the PDM for confirmation through the Data Query process.

### **5.3 Ongoing Study Management Support**

Ongoing study management support will be provided through the use of a Subject Control Report (SCR). The SCR lists all patient ID's by site and includes columns for every expected source data. During data processing, the SCR is constructed during the merge of data sources in such a way so that a “1” is printed if the data source is present for a particular subject and “0” is printed if the data source is absent. The 1's and 0's enable counts of available source data and provide quick and readily apparent flags for data expected but not yet integrated into the merged analysis database. The SCR will be e-mailed to study investigators on a routine and regular basis in order to communicate the status of receipt of data by the Data Coordinating Center.

## **6. Clinical Study Report Documentation**

All results included in final analysis reports will have source documentation from one of three sources.

- SAS to Excel Updateable Tables
- Proc Report Listings
- SAS Listing files

### **6.1 SAS to Excel Updateable Tables**

SAS to Excel Updateable Tables are will be created by SAS programs written with the same name as the Excel table. Running of the program should automatically update the Excel table using the DDE environment.

### **6.2 Proc Report Listings**

SAS Proc Report listings are used to generate tables with patient listings. They are designed to be copied and pasted directly into a clinical study report.

### **6.3 SAS Listings**

All results not generated by SAS to Excel tables or SAS Proc Report Listings will be included in a SAS listing file.

## **7. Backup procedures**

The Primary Data Manager is responsible for maintaining backups of all current data bases. This may be on CD-R/CD-RW, Zip drive, or tape. Backups should be made whenever substantial work has been performed at the analyst's discretion. A remote backup of all raw data (Access and electronic) files will be performed by copying all files to a secure University server on a weekly basis.

## **8. Forms Appendix**

### **8.1 CRF Data Query Form**

