Chapter 21

Data Management

Identifying Subjects: Each person contacted by the study will be assigned a randomized 8-digit identifier through the data entry system. This 8-digit number will represent the person on all forms. To reduce the possibility of confusion, the first page of each form should have a label attached containing a code39 barcode, and an acrostic that represents the first 3 letters of the person's last name followed by the first 2 letters of their first name. For example, John Smith acrostics would be SMIJO.

Duplicate IDs: Prospective LLFS Participants are being identified from many different sources. As a result, occasionally duplicate IDs for the same person will be generated. When Field Center Personnel discover a duplicate LLFS ID, they should complete the "LLFS Fuzzy Match Form" (*Please refer to Appendix 1 to this Chapter*). Based on the information obtained from this form, the DMCC will generate a "MATCH" dataset; this will contain the correspondence between the respective LLFS ID and a SUBJECT. A *SUBJECT* will be created in order to uniquely define people who have two or more LLFS IDs.

In addition to this process, a specially written program is run at regular intervals by DMCC staff to identify other duplicates. The output of this program is then sent to the appropriate Field Center for review, i.e. to confirm or deny the matches. Once confirmed, the matches will be added to the "MATCH" data set.

Clinic And Recruitment Staff Identifiers: Every person involved in contact with subjects or prospective subjects will have a 3-digit identifier. This identifier should be present on all data collection forms to identify the person performing the test, administering the interview, or the person responsible for verifying that subjects have completed materials that were not administered by study staff.   The master list will be maintained through the data entry system by the field center manager or a designated alternate.

Field Center Codes: Field center identifiers should be circled on the first page of each form. They are not entered into the data entry system, as the data entry system will automatically keep track.

Obtaining Study Forms: Forms may be downloaded as PDF files from the study web site at http://dsgweb.wustl.edu/llfs/.

Data Conventions – General Rules for Completing Forms: The following is a list of all the data conventions for the study when filling out forms.

1. Entering Dates: All dates should be entered as the day of the month, followed by the three-letter month abbreviation in the local language, and a four-digit-year. As an example, October 22, 2005 would be entered as "22OCT2005" in English, or "22OKT2005" in Danish. All year values should be entered as 4-digit years.
2. Block capital lettering should be used to fill out alphanumeric data fields. Where check boxes are indicated, a solid, easily-legible "X" should fill the center of the box.
3. Leading zeros are not necessary on the form, or in the data entry system, e.g., if an age field has three spaces, and the person's age is less than 100, it is not necessary to enter 085 for 85.
4. Coding Missing data items and/or Missing Forms: Missing items should be coded according to the rules given below for any questions that are asked and not answered. Questions that are not asked due to a skip pattern within a form are the only fields on any given form that should be left completely blank. If a form is not filled out at all, then the date should be coded according to the rules below.

    R: *Refused*: Used if subject refuses: to answer a question, fill out a form, or perform a procedure when there is no specified method of handling it in the protocol. If the subject refuses a questionnaire or procedure in its entirety, enter "R" in the date field for that questionnaire or procedure.

    N: *Not applicable*: Used when subject is unable to perform a procedure, and there is no specified method of handling it in the protocol or form.

    P: *Missing due to Protocol*: The procedure/question is not relevant (e.g., gender-specific questions, or subject was excluded from answering the question, or performing the test due to a physical condition, and there is no specified method of handling it in the protocol or form.)

    U: *Unavailable*: Used if a measurement/question is not completed for any reason other than the above three, and there is no specified method of handling it in the protocol or form.

    D: *Don't Know*: Used sparingly, after all probes have failed. May not be available for some forced choice questions.

5. <u>How to correct errors</u>: Corrections are made by drawing a single line through the incorrect entry such that the original entry is readable and entering the correct data above the incorrect entry. All corrections must be initialed and dated and made only by appropriate site staff (e.g. study coordinator, PI, sub-investigator, pharmacist, neuropsychological examiner) Do <u>not</u> write over the previous entry. Erasures, liquid paper (white-out), correction tape, etc. are <u>not</u> permitted.

The Data Entry System: The LLFS data entry system consists of a SAS$^®$ System based application written for the LINUX platform. The application is accessed using an X-Windows interface through a virtual private network (secure and encrypted) connection to the LINUX machine located at Washington University from a personal computer. Hardcopy output is printed locally at the field center from PDF files produced on the LINUX server. These PDF files are obtained from a web page designated for the field center's use. It is designed to use menus and "point-and-click" for ease of navigation, while still providing the traditional keyboard method of data entry. Since data entry takes place on the server at Washington University, there is no need for data transmission.

Each field center has a PC provided by the Data Coordinating Center to assist in data collection. However, the required software to access the data entry server can be installed on any PC equipped with a Pentium-level processor and Ethernet card. Each field center may install the software on as many machines as they need. Multiple users can use the data entry system at the same time.

The LLFS data entry system is designed to ensure data security and improve its integrity, while providing easy access and use. It is also built to avoid making users learn programming, or operating system commands beyond a few common ones. The menu system and programs should handle the bulk of the system interfacing necessary for the entry, storage, and management of the LLFS data. There are programs running in the background to do ID look-up, field validation, cross-field validity checking, and computations. These processes are employed to improve the data integrity. However, they also consume computer resources, so you may experience occasional delays when you are in the data entry system. Because the data entry system is accessed through the Internet, response time may seem slower than on your local PC.

System Access And Security: To gain access to the LLFS data entry system, each user is required to have a valid userid and a pass phrase for that userid. The userids are assigned by the field center's data manager, along with an initial pass phrase. Each individual user should have his or her own userid. Each user should change their initial pass phrase the first time that they log on to the data entry system. The pass phrase should be something easy to remember, as it should not be written down so that other people may see it and use it to gain access to the data and programs. When you have finished working on the system, please exit the Main Menu and return to the "Login" screen. If you have finished for the day, press the "Exit LLFS Data Entry System" button. This will close both the data entry system and the login to the LLFS server at

the DCC. This will prevent unauthorized people from having access to the system. The system also keeps track of who has been using it, and this should be as accurate as possible. If you exit the main menu, the next person to sign on will have to use his or her own userid.

In addition, there is a second level of security within the menu system. The LLFS data entry system uses the concept of "authorization levels" for each userid. This means that certain functions and menus are only accessible if the current user is authorized for that level of access. If someone attempts to access a function for which he/she has insufficient authorization, an "ACCESS VIOLATION" message will appear on the computer monitor. Should you feel that you should have access to a particular function that you are not authorized for, discuss it with the data manager at your field center.

The following documentation should give all the information necessary to use the system, its menus and submenus, data entry screens, and support and management functions. As with any piece of documentation, it is constantly undergoing revision. Should you find any errors, or would like clearer instructions on how to do a task, please contact the Data Coordinating Center.

Starting the LLFS Data Entry System and Logging In: Once you have logged into the LLFS server at the DCC, you may not see anything happen for several seconds. If nothing happens after a few minutes, call the DCC. The apparent delay is for the loading and initialization of the SAS System®, and for the startup of the LLFS data entry system. During this initialization period, the logo for the SAS System should appear on the screen. Once the loading has finished, the "Login" screen will appear. This screen has room for two fields: userid and pass phrase, along with a pushbutton to close the data entry system. Enter your userid, and then your pass phrase. The pass phrase must be at least eight characters long, with at least one non-alphabetic character. Spaces are allowed, and the pass phrase is case-sensitive. The pass phrase will NOT appear as you type it. Should you type either the userid or the pass phrase incorrectly, the system will not allow you to log in, but will notify you and ask you to re-enter them. You are allowed up to three tries before the data entry system automatically closes and ends your session. Both the userid and pass phrase are case sensitive.

Once all of the form screens have been finalized and the data entry system completed, a comprehensive data entry instruction manual will be provided to each Field Center's Data Manager.

Chapter 21: Appendix 1

| LONG LIFE FAMILY STUDY | (Affix Label Here)

Participant ID: _____

Participant Name Code: _____ | Date Form Filled Out:
☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
d   d   M   M   M   y   y   y   y
(e.g., 10JUN2005)

Interviewer Code: ☐ ☐ ☐

Circle Field Center Location:
BU   CU   DK   UP |

LLFS "FUZZY MATCH" FORM

As of 11/29/07, a New, Updated Version of this Form is In-Progress

Date(s) of Submission: ____ ____ / ____ ____ / ____ ____ ____ ____

Contact Person:_____

CONTENTS:

PARTICIPANT ID MATCHING: