SEARCH MOP - Section 6
Data Management
Table of Contents

6. DATA MANAGEMENT 6-1

6.1. OVERVIEW 6-1

6.2. DATABASE MANAGEMENT SYSTEM 6-1

6.2.1. Local Tracking Application 6-1

6.2.2. Central Web Application (SEARCH Website) 6-1

6.3. GUIDELINES FOR DATA ENTRY 6-3

6.3.1. Participant and Form Selection 6-4

6.3.2. Participant Initialization 6-4

6.3.3. Data Entry of SEARCH 3 Consent Form 6-5

6.3.4. Data Validation Procedures 6-6

6.3.5. Edit Checks 6-7

6.3.6. Reporting 6-7

6.4. DBMS SOFTWARE UPDATES 6-7

6.5. ELECTRONIC TRANSFER OF CENTRAL LABORATORY DATA 6-7

6.6. DATA CONVERSION AND EXTRACTION 6-8

6.7. DATABASE CLOSURE AND DOCUMENTATION 6-8

6.8. SECURITY 6-8

6.9. DISASTER RECOVERY 6-9

6. Data Management

6.1. OVERVIEW

The data management system used for data collection during the SEARCH study utilizes a combination of a web browser-based interface and a local tracking application, with most electronic data stored and managed centrally at the Coordinating Center (CoC). Some information regarding the tracking of site participants will be kept locally at the site. User-friendly screens, matching case report forms (CRFs), were developed using ColdFusion supported by a SQL Server database on the backend. All participant data entered by clinic staff resides in the SQL database on a server located securely on the internal network.

6.2. DATABASE MANAGEMENT SYSTEM

The **SEARCH Data Base Management System** is divided into two systems: 1) a local tracking application on the site's PC, and 2) a centralized web-based application from the CoC.

6.2.1. *Local Tracking Application*

The local tracking application allows a User to track and manage participant contact information, record visit data including dates and types of contact and assigned ancillary studies, and produce a variety of reports and reminder schedules. This system tracks case information and upload critical data to the central database system. Each center must use a site specific tracking data base to perform these functions. A Microsoft Access tracking database developed by the Coordinating Center is described in Section 7.

- The local tracking application allows sites to manage User contact by permitting entry of contact information such as visit type, location, date, ancillary study and other required information.

- The local tracking application has been installed locally at each site. A User begins the application by double-clicking the icon located on the PC's desktop. Once the application starts, a User will be required to enter a Username and password. Each User will be provided their own personal login information. Once a User has been properly authenticated, an Initial Application Menu will appear from which they will choose to initialize a new participant or to edit an existing participant.

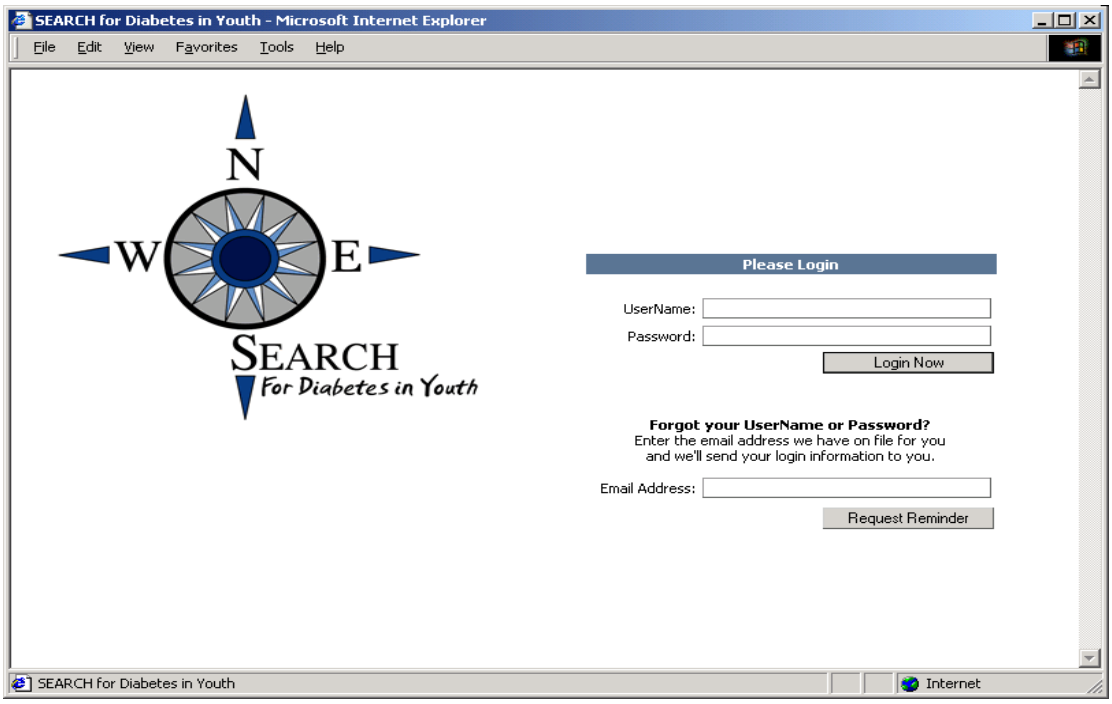
6.2.2. *Central Web Application (SEARCH Website)*

The central web application allows a User to interact with all phases of the clinical data. Data Entry screens for each of the study forms are available in this application. Users are

allowed to enter and edit data, view reports, print schedules, and more based on security access.

The central web application is used to enter the majority of the data collected during the study. To access the application, a User should double click the Internet Explorer icon on the desktop. Once a web browser is invoked on the client computer, the following URL should be entered:

https://search.phs.wfubmc.edu/



A Username and password is provided for each person in the study by the Coordinating Center at the start of the study. The site should contact Ken Wilson at the CoC when setting up study staff members or student account; likewise, when deactivating an account. (11/11) Passwords may be changed at any time throughout the study by simply clicking the ***Profile*** menu item and then changing the password field. The initial page of the website contains Username and Password fields for logging into the website. The system is password-protected to prevent unauthorized access. Once a valid Username and password combination are entered and verified, a User can click the **login** button below these fields and proceed to the main screen.

If a User has forgotten either their Username or password, their email address can be entered in the bottom field labeled **email address**. Using "**email address,"** the Username and password associated with that email address can be found and transmitted, by email, to that address with the login credentials. Data encryption techniques during

transmission have been employed throughout the web application to further enhance security and meet HIPAA and 21CFR11 regulatory guidelines.

6.3. GUIDELINES FOR DATA ENTRY

As a User works with the database management system, they should be familiar with the following operations:

- Each screen displays a certain number of fields or slots where data from SEARCH forms may be entered. These fields may appear on the screen as boxes varying in color from the screen background. Other fields called "radio buttons", "list boxes", and "check boxes," allow a User to select from several choices. If information has been omitted from a form, it may need to be specified as "missing".

- A User can move from field to field by using the *Tab* key, or by clicking on another field.

- If a mistake is made entering data while in a specific field, backspace over the error and retype it. If the mistake continues after moving to a new field, use the mouse, Shift-Tab or Tab keys to move to the needed field.

- In some cases the system will perform necessary calculations previously performed by clinic staff. The cursor will not move to those fields. A User must check the system's calculations against those made by hand. If discrepancies appear, first check to be certain data have been entered correctly. Report any unresolved problems to the Clinic Coordinator.

- When the end of a screen is reached, check all entries to assure correctness. Save each form after entering data. If all form data has not been entered, a User can click the *Save Incomplete Form* button saving the form "as-is" for completion at a later date/time. If all data has been entered, a User should click the *Finalize Form* button saving the form and indicate that data entry has been completed.

- Each time a participant ID is entered, the system determines the validity of the ID. If the ID is not valid, form access is denied and the User is notified of the problem. Participant ID's are not available for data entry until they have been uploaded to the website via the Tracking Database.

- Missing data fields should be assigned a **[-9]** with the exception of fields where a year is to be entered. Missing [year] data fields should be assigned **[1800]**.

- If there is invalid data on a paper form, record it as missing on the computer screen and consult the Clinic Coordinator. Be sure to maintain a paper log of these discrepancies,

recording the participant ID, visit form, date on the form, today's date and where the problem occurred.

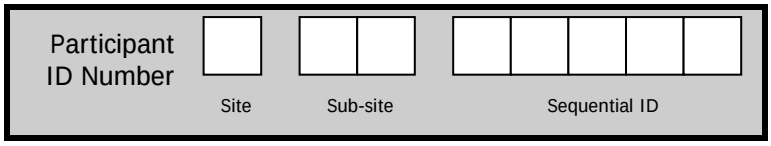
- Sites always have the ability to correct forms that may have been entered incorrectly. Additionally, the Project Managers have the ability to delete /restore forms that are entered incorrectly. When making the corrections, be sure to select the appropriate visit before selecting the form to edit. (11/11)

6.3.1. *Participant and Form Selection*

A User may enter the participant ID once access to the Data Entry portion of the system is obtained. Similar to the username-password combination on the log-on screen, no further access is given until a valid participant ID is entered. This method prevents data entry error in the selection of the participant and ensures that all data is entered to the correct participant. After a correct ID is given, a User is presented with a summary page of information that is known about this person. From this point, a User can access a variety of reports and data entry screens for this participant. Forms may be accessed by clicking on the name of the form in this detail window.

6.3.2. *Participant Initialization*

The SEARCH participant ID is used on all data collection instruments and labels for identifying items associated with individual study participants such as laboratory specimens, supporting documentation, etc. A place is provided on each form to write a participant's SEARCH ID using the following format:



All study reports that refer to individual participants use this numeric ID as the only means of identification. On reports and study printouts, the participant ID number is printed without separating dashes or spaces (for example, a valid study ID that appears on a report will be of the form "30000042"). Following is a description of the convention used for generating SEARCH participant IDs:

ID Element	Description	Valid Values
Site number	A one-digit numeric code for the SEARCH clinical site where the participant is enrolled. Sites have been identified sequentially, beginning with South Carolina and moving east to west.	Seattle - "5" Los Angeles - "4" Colorado - "3" Cincinnati - "2" South Carolina - "1"
Sub-site number	A two-digit numeric code to identify sub-sites for those SEARCH clinical sites where distinction between enrollment sources is necessary.	00 - 99
Sequential ID number	A five-digit sequential number that uniquely identifies (within the SEARCH clinical site) each participant registered.	Integer values with leading zeros, ranging from "00001" to "99999".

Other identifying numbers found on any potential data sources (e.g., hospital or clinic chart numbers, social security numbers, insurance identification numbers, etc.) ***should not be used as the SEARCH study ID***.

6.3.3. *Data Entry of SEARCH 3 Consent Form*

A web-based form will be data-entered for every Phase 3 consent form that is signed. In some rare instances, information may change on a participant's hardcopy Phase 3 consent form. A corresponding change may be needed on the data-entered version to keep the two consistent with one another.

- If re-consent was not needed for a visit, and a new hardcopy form wasn't signed, then nothing new should be data entered.
- If a new hardcopy Phase 3 consent form is signed, then the information should be data-entered as a new consent form on the web.
- If a participant's response on a hardcopy Phase 3 consent form is changed, then an edit should be made to the corresponding web-based consent form.
- The electronic version of the consent form should reflect the information within the signed form. The following guidelines may be used when completing the response for each of the following: saving of blood, saving of urine, and saving of DNA:

I. If the signed consent form indicates (either explicitly or implicitly) that the participant has agreed to sample storage on that date, then 'Yes, consent given' should be data-entered.

II. If the signed consent form indicates (explicitly) that the participant has refused sample storage on that date, then 'No, consent refused' should be data-entered.

III. If the question is neither explicitly nor implicitly asked on the signed form, then 'Question not on form' should be data-entered. This includes signed forms on which the relevant section has been explicitly marked as 'not applicable.'

IV. If the question is explicitly asked on the signed form, but no response to the question is recorded by the signee, then the response should be '-9 - Known missing' on the electronic version. Signed forms on which the relevant question is explicitly marked as 'not applicable' should NOT be data-entered as '-9 - Known missing.' See III above.

- Enter the date on which the consent form was signed. If a change is made on a hardcopy consent form, and the form is re-signed, then the date entered should be the date the form was re-signed.

- Each study center may determine whether or not to complete the consent form version field, as well as the specific information to be entered in this field.

6.3.4. *Data Validation Procedures*

The database management system performs several validation checks during the entry process. Data must 1) match the correct type (numeric data in numeric fields), 2) be in the correct range of valid responses, and 3) be appropriately marked when missing. Data that fail the established validation checks generate messages or prompts that describe the problem and required actions.

All form questions are pre-assigned missing values, e.g., null, for the purpose of data entry. Data entry screens require a set degree of completeness before a form can be accepted as 'final.' For incomplete forms, the missing value is entered into the database. Validation checks will be applied during the data entry process. Insofar as possible, checks will be programmed using JavaScript routines and made as the clinic staff enter data from each CRF.

A related issue is data out of "expected" and "valid" ranges. For example, an 8 year-old boy should have a systolic blood pressure between 62 & 124 mm/Hg. You may encounter valid values outside the expected range, e.g., parental age > 110 years. In such cases the system will notify you that the data entered is out of *expected* range but entry **is** allowed. A query list will be posted asking you to verify data values outside the expected range. Once data has been validated/verified, the posted query will be satisfied and removed from the form's query list.

6.3.5. *Edit Checks*

Computerized data validation routines will be used to enhance data quality, including,: a) initial screening of data, using logic and range checks built into data entry screens; b) cross-form functional and consistency checks; and c) edits assessing the serial integrity of data, particularly in longitudinal studies.

6.3.6. *Reporting*

Monitoring study data will occur at both the Coordinating Center and the site in order to achieve and maintain a high level of quality. Some of the monitoring and quality control reports will be transmitted to the sites for immediate action and attention; other quality control and monitoring reports will be generated for the Steering Committee.

6.4. DBMS SOFTWARE UPDATES

Periodically, during the SEARCH study, the Coordinating Center will update the database management system. If significant changes are made to the system, clinic staff will be informed via the Main page of the web site in order to note the new changes.

When updates are needed on the local tracking application, the sites will be notified that they need to download an update executable from the web site and they will be provided instructions as to how to complete the update. During any local application updates, a backup of the current database will be made to a specific directory on the site's PC.

6.5. ELECTRONIC TRANSFER OF CENTRAL LABORATORY DATA

Data from the Central Laboratory is delivered to the Coordinating Center via email and uploaded to a repository on the server. Specific import routines have been developed to verify and merge these data with the main database.

6.6. DATA CONVERSION AND EXTRACTION

SAS analysis files can be extracted from the database using SAS/Access. Programmers continually develop routines to create other specialized analysis files from the SQL Server database or the SAS database. Prior to merging or extracting any data into or from the database, merge/extraction routines are being developed and will be thoroughly tested. Since data arrives from differing locations, verification includes consistency checks across all platforms as well as any other routine checks. All routines are properly documented and changes and updates to the code noted.

6.7. DATABASE CLOSURE AND DOCUMENTATION

Upon study completion, after all clinic and laboratory data have been collected and filtered through various quality control routines, the resulting SQL Server database will be converted to SAS and ASCII data sets and certified. The database will be taken offline and archived on magnetic tape and/or CDROM. The final data sets will be certified and issued version numbers to synchronize analytic efforts and will be distributed in SEARCH in compliance with Steering Committee and institutional policy. The choice of media on which to copy and distribute copies of the database to the investigators will depend upon the systems and the media available at that point in time.

Documentation will be prepared that contains a brief overview of the project, the goals, and the type of data collected. This will be followed by a list of variable names, their positions, and short descriptions of each variable contained on the media.

6.8. SECURITY

Data are normally transmitted across the Internet as plain text. It is possible, but highly unlikely, for someone to monitor this traffic and, using the proper equipment, reconstruct the individual pieces into the original data. Because of this threat, we employ a digital server certificate. This certificate allows the communications between the web server and the client system to be encrypted. This encryption is as advanced as is currently allowable by the United States Government. This mechanism is the same as is used by the banking industry and for electronic commerce. We feel strongly that this system will provide more than adequate security against unauthorized use.

Restricted areas of the web site are protected by User login. Prior to gaining access to the restricted area, a User is required to enter a Username and password that will be checked against a database. If the combination is correct, a “flag” will be set to allow the User to enter certain areas of the web site. For security purposes, once a User has successfully logged into the system, inactivity for a period of 2 hours will automatically force the User to

re-authenticate prior to using the system again. We strongly recommend that Users log out of the system before leaving their work area for any extended period.

WFUSM is protected by a Cisco firewall that limits the source and type of traffic coming into the institution. This product remains under constant monitoring and control.

6.9. DISASTER RECOVERY

All data, programs, code, documents, etc. associated with the SEARCH project will be backed up nightly to a DLT tape library. These tapes are kept indefinitely and are located in a fireproof cabinet that remains locked at all times. Periodically, copies of tapes are moved to an off-site location for storage. In the event that there is any loss of data, the information can be restored from tape in a matter of hours. The entire PHS computer facility is provided with conditioned power, UPS capability and environmental sensors with notification protocols.