SEARCH 4 MOP
Section 4 - Data Management
Table of Contents

4. Data Management

4.1. OVERVIEW

The data management system used for data collection during the SEARCH study utilizes a combination of a web browser-based interface for clinic use and a web browser-based interface for participant direct data entry of cohort surveys, 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 participants and clinic staff resides in the SQL database on a server located securely on the internal network.

4.2. DATABASE MANAGEMENT SYSTEM

The **SEARCH Data Base Management System** is divided into three systems: 1) a local tracking application on the site's PC; 2) a centralized web-based application for clinic staff to enter questionnaires and study management; and 3) a participant direct data entry web-based application where survey only participants can directly enter data.

4.2.1. *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 uploads 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 5.

- 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.

4.2.2. *Central Web Application for clinic staff (SEARCH Website)*

The central web application features a 'Data Management' section where users can enter, review and edit data, resolve outstanding queries, view and edit participant contact, IPS

and IPV statuses and upload data. Also featured is a 'Dashboard' section where users can access directories, committees, ancillaries, a wide variety of study resources, papers and presentations, reports and more based on security access.

To access the application, a User should open their web browser go to the following URL:

https://www.searchfordiabetes.org
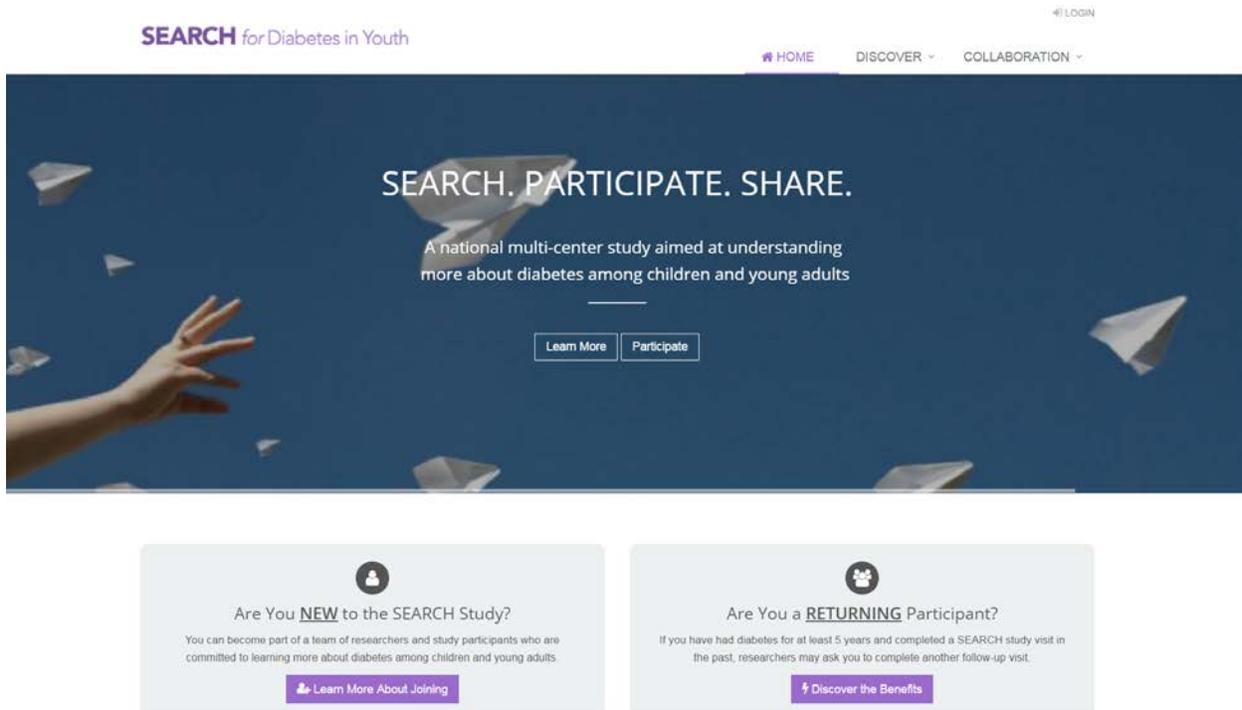
A Username and password is provided for each person in the study by the Coordinating Center at the start of the study. The sites should contact the CoC when setting up study staff members or student account; likewise, when deactivating an account. Passwords may be changed at any time throughout the study 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 their password, their email address can be entered after clicking the 'Forgot Password' link at the bottom of the login page. An email will be sent to the user that contains a link that they can use to reset their password.

4.3. GUIDELINES FOR DATA ENTRY

To access the data entry area, click on the 'Data Management' menu item, then 'Clinic Data Entry.' This will display a search box to enter a participant ID. Once an ID is entered/selected, click 'Go' and the 'Participant Hub' is displayed. The 'Participant Hub' lists SEARCH 4 Clinic data entry forms as well as SEARCH 4 Survey Packets for participants and parents, as applicable. There is also a link to the legacy SEARCH data entry website to review SEARCH 2 and SEARCH 3 forms. In addition to listing the forms, mode of entry, status, date completed and updated, number of queries and used status are displayed.

For the Participant and Parent Survey Packet sections, if a participant/parent has not yet been sent the link to their survey packet, the clinic staff person will first need to assign an age-specific form packet by choosing the correct age and gender category from the 'Select Form Packet' dropdown list and click submit. Once 'Submit' is clicked, a list of age/gender specific modules will be displayed. If the participant/parent has already entered their data, clinic staff can review the data, resolve any outstanding queries and finalize the forms.

For the Clinic data entry forms, once a data entry form is selected, it will display with the first field 'Re-enter Participant Identification.' This field and the footer fields ('Form Used,' 'Completed Staff,' 'Reviewer Staff,' 'Data Entry Staff,' 'Mode of Administration,' 'Date Completed (Visit Date),' 'Date Reviewed,' 'Date Entered') are required. All other fields on the forms have validations that are warnings meaning the data saves with warnings and allows resolution of warnings after save.

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," if the option is available.

- 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/form 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 Form in Process* button saving the form "as-is" for completion at a later date/time. If all data has been entered, a User should click the *Save Form Finalized* button saving the form, indicating that data entry has been completed.

- Each time a participant ID is entered, the system determines the validity of the ID. If the ID does not match the ID on the participant hub, the data will not save.

- If there is invalid data on a paper form, record it as missing on the computer screen, if the option is available, and consult the Clinic Coordinator. Be sure to maintain a paper log of these discrepancies, recording the participant ID, visit, 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 review and verify participant data entered forms.

4.3.1. *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.

4.3.2. *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.

4.3.3. *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.

4.4. 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.

4.5. 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.

4.6. 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.

4.7. 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.

4.8. 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.

4.9. DISASTER RECOVERY

All data, programs, code, documents, etc. associated with the SEARCH project will be backed up nightly using the Medical Center's Enterprise backup system. In the event that there is any loss of data, the information can be restored in a matter of hours. The entire PHS computer facility is provided with conditioned power, UPS capability and environmental sensors with notification protocols.