SOL Youth
Manual 05
Data Management

August 22, 2012 - Version 3

Study website - http://www.cscc.unc.edu/hchs/

SOL Youth Manual 5 - Data Management

Contents

1. SOL YOUTH Field Center Data Management

This data management manual focuses on SOL Youth data collection and processing procedures at the field centers, at the central agencies and at the Coordinating Center (CC.) Topics include identification of participants, data collection forms and procedures, instructions for completion of paper and electronic forms (including corrections), data transfer and naming conventions, revisions to manuals and forms, and general guidelines to ensure data security, accuracy, and accessibility at field centers and central agencies.

1.1. Web-Based Data Management

The SOL Youth data management software Carolina Data Acquisition and Reporting Tool (CDART) is a system created to manage data collected in the SOL Youth study. The field centers and reading centers will be using this system to collect SOL Youth data. CDART runs over the Internet from any computer with a high speed internet connection, using Firefox or Google-Chrome web browser (do not use Internet Explorer.) By design, CDART supports data entry either interactively "on screen" during data collection, or from paper forms. However with the exceptions noted below, the data collection protocol specifies direct entry into the CDART as the mode of data collection. Work station configuration at the field centers should be arranged accordingly. In all circumstances when information is collected on paper form, transcription into the CDART should occur before the participant leaves the field center, to allow for clarification and collection of missing items. The forms that may be collected on paper first are:

Recruitment/Administrative

- Household Screening Roster Form (HSR)
- Screening Call Tracking Form (SCT)
- Individual Eligibility Checklist Form (ELE)
- Participant Safety Screening Safety Form (PSE)

Self-Administered

- CDE - Child Depression Inventory
- MASC-10 (MAE)
- Pubertal Development (PDE)
- Tobacco Susceptibility (TUE)
- Alcohol Susceptibility (AUE)
- How I feel scale (LIE)

The separate CDART User's Guide provides specific instructions on using CDART. Usernames and passwords for the system are provided by the CC in a secure and confidential manner to each site's project and data managers.

CDART provides several major functions that are fundamental for data management:

Data Entry: Allows data collection of forms to be keyed, edited and updated.

Reports: Provides customized reporting based on study need. The CDART will generate participant lists, form inventories, bar-coded form labels, etc. to help the field centers with any data management-related tasks. Requests for reports or lists not provided by the CDART can be made by the field centers to the CC. Participant recruitment, scheduling, and other managing of participant flow is the responsibility of the field center.

Data Transfer: Allows field center and/or central agency data to be sent to the SOL Youth Coordinating Center for inclusion into the consolidated study database.

1.2. CDART Data Entry URL

https://cdart.cscc.unc.edu/SYOUTH

Copy and paste the link into the address bar in your Firefox or Google-Chrome web browser. Remember; **DO NOT** use Internet Explorer (IE.)

1.3. SOL Youth Participant ID Numbers

Study ID numbers for participants enrolled in the SOL Youth study are created and assigned to each field center by the CC as part of the recruitment process (see section "Eligibility Criteria", of manual 1 for more details.) Bar coded study ID labels for use on data collection forms can be generated using the CDART by selecting the study ID for which you want to print a sheet of label stock.

SOL Youth participant ID's are 8 characters long with the following format:

- Character 1: Site Identifier (B, C, M, S for Bronx, Chicago, Miami, San Diego, respectively)
- Characters 2-7: Participant ID number
- Character 8: Check Digit (based on published algorithm, used for input validation of ID)

1.4. Identification Information on Data Collection Forms

The information that identifies each form as a unique record in the CDART is the key field information contained in the "header" box at the top of the first page on all forms (see example below.) The following guidelines should be observed in filling out the "header" information located at the top of the first page on all forms.

SOL Youth Personal Information Questionnaire

ID NUMBER:									FORM CODE: AEE VERSION: 1 1/10/12	Contact Occasion	0	1	SEQ #		

When paper forms are used, preprinted, bar-coded participant ID labels should be used in the header of the form whenever possible. Study ID labels can be generated using the CDART "Report" interface. Detail information on "Reports" can be found in the CDART manual. If the 8-digit ID number is handwritten, care must be taken to make sure the number is accurate and legible for data entry.

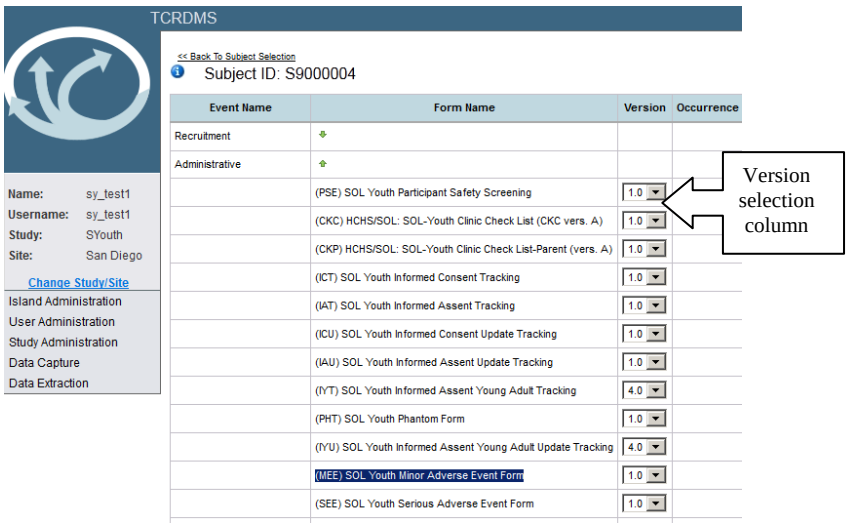
1.4.1. Form Code and Version

The form code is unique for each data collection form in the SOL Youth study. Form codes and versions are preprinted on all forms. If form content changes during the study, those changes will result in the version being updated from the initial version, "1", to the next whole number.

It is the sites responsibility to ensure that the appropriate versions of the forms are being used at all times. Note that the CDART will load the current version of a form automatically. The sites will be informed via email of any form changes. The new version of the paper form will be uploaded to the study website.

1.4.1.1. Selection of Form Version in CDART

Sometimes form versions have to be selected during data entry when the current version is not the version being keyed in. When different versions are available CDART will provide a column titled Version that will allow form version selection. The screen looks as follows:



1.4.2. Contact Occasion

The Contact number will be used to differentiate a participant 'contact occasion' or visit over time. The contact occasion is pre-printed on forms. Given that this is the first contact time for data collection for SOL Youth participants, all interview forms and procedures are collected at contact occasion "01." As the study participant follow-up schedule is defined over time, contact occasion will be incremented.

1.4.3. Sequence Number

The sequence number (Seq #) enumerates and distinguishes multiple forms collected during a specific (i.e. the same) visit number. For all forms at the regularly scheduled contact the sequence number is simply "01." Where appropriate, the sequence number has been pre-coded on the form. Any forms that are repeated for use in quality control purposes will have a different sequence number (such as "02", "03") to distinguish them from regularly scheduled contact forms.

1.5. Administrative Information

1.5.1. Code Number IDs

Certified site staff will be assigned a 3-digit code number by the CC. The staff member must use this number on data collection screens that ask for a staff code number ID. This number is assigned by the CC. Contact the CC to obtain numbers for new staff members.

1.5.2. Date

The date to be recorded onto a data collection form header is the date of the participant contact or specimen collection (i.e. clinic visit examination date), or the date the form is completed. The date must conform to the month / day / year format as specified and be within the bounds of the time line. Pre-dating and post-dating of forms should not be done.

1.6. Data Collection

1.6.1. Background

SOL Youth uses a combination of data collection methods: direct data entry, recording on paper forms followed by data entry, and forms collected on paper only (with no data entry.) The purpose of this section is to provide instructions for completing forms. Prior to working with the forms, both this section and the specific question-by-question set of instructions for each form (QxQ) should be read carefully. (The QxQ instructions will follow the paper form in the corresponding MOP. They are also posted in the study website, following corresponding paper form.)

1.6.2. Form Structure

The CDART data entry screens are designed to parallel the data collection paper forms. The general layout of the paper forms is as follows:

First Page of Form:

- Form Title
- "Header" Information
- Participant's ID Number
- Form Code and Version
- Contact Occasion
- Sequence Number
- Data collection questions

In the CDART, the Form code, contact occasion, and version are displayed in the left-pane of the form entry screen. The form name is displayed in the heading of the form entry screen and in the form grid.

Form entry Screen

Form Grid

Event Name	Form Name
Recruitment	
	(SCT) SOL Youth Screening Call Tracking Form
Administrative	
	(ICT) SOL Youth Informed Consent Tracking
	(IAT) SOL Youth Informed Assent Tracking
	(ICU) SOL Youth Informed Consent Update Tracking
	(IAU) SOL Youth Informed Assent Update Tracking
Procedures	
	(ANT) SOL Youth Anthropometry
	(SBP) SOL-Youth Sitting Blood Pressure
	(BIO) BIOSPECIMEN COLLECTION FORM
	(FST) SOL-Youth Fitness Step Test
	(CHM) Chemistry Import
	ACT
Questionnaires-Child	
	(DCE) SOL-Youth Demographics - Child
	(CDE) SOL-Youth Child Depression Inventory - Child
	(TUE) SOL-Youth Tobacco Susceptibility - Child
	(AUE) SOL-Youth Alcohol Susceptibility - Child
	(LSE) SOL-Youth How I Feel Scale - Child
Questionnaires-About Child	
	(MHE) SOL-Youth Medical History – Parent (each Child)

1.6.3. Recording Responses to Questions

Many of the questions in the SOL Youth forms have a set of pre-coded responses, with instructions to "enter the appropriate response" (code) or "check all that apply" (checkbox.) However, a few questions require a written response. Some questions request a textual response. Others request elaboration of an "other" or "specify" response from a previous question. Space is provided on the form for those unstructured written responses.

If a participant's answer does not logically fit into one of the pre-coded answers, the interviewer must specify the response by recording it on the form beside the pre-coded answers. Data entry personnel are trained to enter the additional data into note logs.

The data collection practices below must be followed at all times to assure that the recorded response accurately reflects the participant's answers and that questionnaire data can be converted to a computer-readable format.

Guidelines for the interviewer include:

- Listen to what the participant says and record the appropriate answer if the response satisfies the objective of the question.
- In recording answers to open-ended questions or "other" categories, record the response in the participant's exact words.
- On paper, record in the white space (below the questions) any responses "that don't quite fit" in one of the response categories. The interviewer's notes will help the data analyst to understand points of confusion, difficulty, etc. Notes on paper forms can be entered as "note logs" in the CDART.

On paper, print or write legibly.

- If a participant refuses to answer a question, and "refused" is not a value in the response set, write "refused" in the left-hand margin beside the question and enter **equal signs** ("=") in the response field to signify a double strikethrough.
- A single answer code must be circled / entered for each question to represent the participant's answer.
- A "select all that apply" answer pattern is indicated with a checkbox, or with instructions to "circle all responses that apply".

Some of the questions in the SOL Youth study ask about recall of events over time. The interviewer may assist the participant without violating probing rules by working with him/her on converting dates to duration (e.g. "for how long did you….") or pinpointing dates or events. Another way to help with the collection of more accurate information is to ask the participant to think about the time of year or season when an event occurred.

1.6.4. General Instructions for Completing and Correcting Items on the Forms

General guidelines for the interviewer regarding forms:

Review each form and its instructions prior to use. If you are collecting data on a paper form prior to data entry, verify that you are using the appropriate form by checking its 3 letter form code, version, and date, all located in the lower left-hand corner of the page. Each unique form type will have specific instructions for filling out that form in the Field Center Procedures Manual (MOP 1.) Be familiar with the instructions described in MOP 1 before attempting to complete a form. Print all text responses legibly; do not use cursive writing if collecting data on paper first.

All items fall into one of three main categories: (1) "fill ins", (2) multiple-choice (circle or check), and (3) qualitative information (comments/short-answer questions.) Techniques for completing each of these types of items, as well as making corrections, are described below. A general rule is to record information only in the spaces provided (except for some error corrections.)

Corrections to paper forms should be made in the following manner:

- Cross out the original response with an 'X' in such a way that it is still legible.
- Write the correct response above or to the side of the original response.
- Date and initial the correct response.
- In cases where numerous corrections were made to the same response, the final corrected response should be circled.
- Major changes should be documented with a brief explanation in the margin.

Corrections to electronic forms are made using the CDART. The CDART records the date and time of the update and the user who makes the change. Many corrections will be made in response to queries sent to the field center from the CC.

Do not attempt to correct errors by using correction fluid or erasers at any time. Data collection forms need to maintain the history of data recorded in the event of an audit. The audit log in the CDART maintains this history for forms which have been entered and subsequently corrected (but does not track paper-only corrections.)

When a response is outside the normal limits or seems contradictory based on other data, confirm the data and, if using paper, write "confirmed" in the margin. This will decrease time-consuming queries and help the data entry staff.

Carefully proofread each page of data for legibility, accuracy, and completeness prior to transferring the form to the data entry staff.

1.6.5. "Fill Ins": Recording Information

"Fill Ins" refers to items where the question is given a defined space for recording the answer. Questions asking for a date, participant study ID number, height, weight, etc. will have a limited amount of space for data entry and usually a require the answer to be inputted in a certain format. In the event that the response contains more characters than there is room for in the space provide by the form, indicate the correct response in the form margin near the original response, and enter the value into a note log in the CDART.

Numeric fields may have a preprinted number of decimal places. In this case, the QxQ instructions will specify the number of decimal places to be recorded. Instructions on how to round values to the expected number of decimal places are found in the QxQ instructions.

When a date is recorded, slashes ("/") are used as the separator characters for month, day, and year. These are usually preprinted in the response field on the paper form but must be entered into the CDART manually. The format to be used to record dates is indicated below the boxes. SOL Youth uses the U.S. order for recording dates (month/day/year.)

SOL Youth usually records time using a 12-hour clock, indicating AM or PM. In most cases, colons (":") are used as the separator character for hours and minutes and are typically preprinted in the response field on the paper form but must be entered for questions where hours and minutes are not separate questions.

1.6.6. "Fill Ins": Correcting Mistakes on Paper Forms

If a number or letter is entered incorrectly, the person making the correction should first mark through the incorrect entry with an "X". Then, he/she should clearly code the correct entry above the original (incorrect) entry and initialize the correction, using his/her 3 initials, and record the date of the correction.

If a mistake is made and corrected, and then it is discovered that the correction is incorrect, make a second correction using the same rules as above.

1.6.7. "Fill Ins": Unknown or Inapplicable Information

If an item of this type (either alphabetic or numeric) does not apply to the participant being interviewed, leave it blank. For example, if the participant does not have an "other" phone number, that item is left blank. Similarly, if the form provides spaces for three measurements, but only two are taken, the third space left blank.

1.6.8. Multiple-choice: Recording Information

In this type of question several choices are given for the answer, each having a corresponding letter and/or word. Once it is decided which answer choice is most appropriate, circle the corresponding letter on the paper form. Always circle one letter/word only. Key the letter or word into the CDART when entering the data.

1.6.9. Multiple-choice: Correcting Mistakes on Paper Forms

If a response is coded incorrectly, mark through the incorrectly coded response with an "X" and circle the correct response. Initialize and date the correction.

1.6.10. Qualitative Information: Recording Information

Some forms contain a substantial amount of qualitative data. These short answer and comment questions will be handled differently than "fill in" type data. If these types of questions are filled out by hand on paper forms, write clearly. When keying these responses, any unreadable answers/comments should be answered to the best of the data entry personnel's ability with a note describing what was illegible.

1.6.11. Skip Patterns

Generally, questions are answered in the order presented, with none omitted.

However, a skip pattern will direct that one or more questions be omitted (skipped) when they are not pertinent to the participant's situation. (For example, if question 1 pertains to gender, and question 2 pertains to pregnancy, a male respondent would be directed to skip question 2.)

Skip patterns occur in some multiple-choice items. This may be indicated on the form by an "If _____, go to question #" statement. If response associated with the skip is selected, the next item to be asked is the one indicated in the "go to" statement.

If a skip pattern response is not selected, proceed to the next item in sequence as usual. Occasionally, a skip pattern will occur in a "fill-in" item (such as, "Other, specify".) If the skip criteria are not met, continue to the next item as usual.

1.6.12. Problem Clarification and Data Queries

The CDART is programmed to automatically query out-of-range values inputted during the data entry process. However there may be a need to send queries from the CC regarding data values within or across forms. All queries will be sent electronically to the data coordinator, participants will be identified by ID number, forms will be identified by header information, data items will be identified by question numbers, the original response will be indicated, and the reason for the query will be described. A cover memo will accompany the data queries describing the problem, with suggestions of ways to resolve the problem and a timeline.

1.6.13. Permanently Missing Forms

In the event that a participant is unable to complete an exam, all forms for the contact (or visit) which were not completed should be coded in the CKP/CKC CDART form as it applies, using the appropriate notelog window option to determine a status of the form. Selections available in the Notolog-Window are: "Missing, Not Applicable, Refused…" CDART does not required that you create an empty form as permanently missing. Indicating the Status on the CKP/CKC is sufficient.

1.6.14. Missing/Unresolvable Items

In the event there is no available answer to key in for a certain item (for any reason such as the participant refused to answer, the answer was not recorded on the paper form, the participant did

not know the answer, etc.) then users should open the notelog dialog box and select the appropriate choice from the drop-down menu located at the bottom.

1.7. Security of Paper Forms

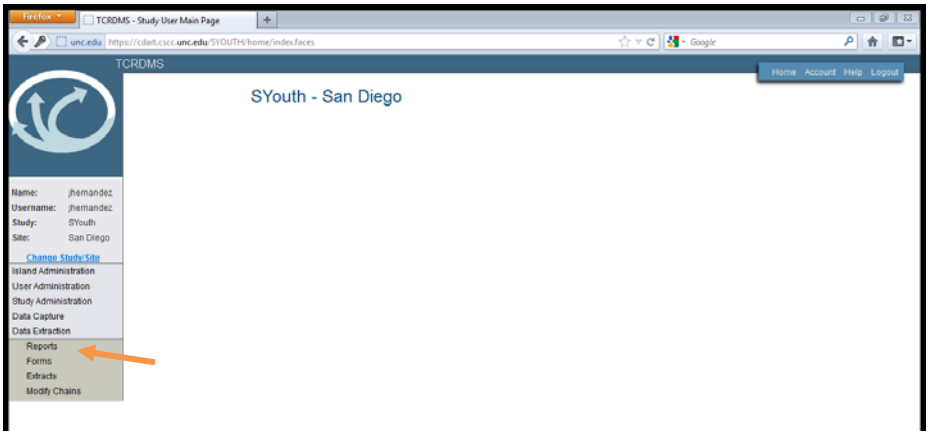
Each clinical site is responsible for assuring that participant study data is stored in a secure location that meets participant confidentiality requirements.

1.8. Data Management Reporting

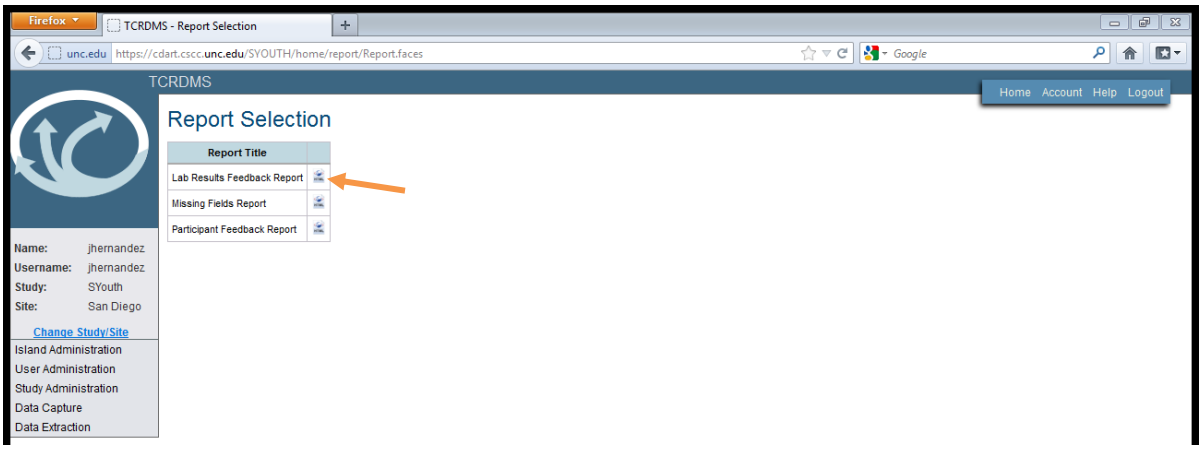
The CDART has numerous reporting programs to facilitate data management at the sites. The SOL Youth CDART User's Guide contains the documentation of the reports available in the CDART. Information on updates and changes to these reports will be provided through the Numbered Memo communication (Section 13.12.1) from the CC to the field centers. As these CDART reports are updated or changed, training conference calls with the field center data coordinators or project managers may also be scheduled.

1.8.1. Lab Results Feedback Report

To generate the lab results feedback report log onto the CDART system and select the appropriate site. Next select "Data Extraction" and then "Reports" from the left-hand side of the screen.



Click on the icon beside "Lab Results Feedback Report" on the next screen.



The following screen requires staff to enter the participant ID. After the ID has been entered, click "Ok". The generated report can be printed and given to the participant. An example of the Lab Results Feedback Report is attached in Appendix 4.2.1.

1.8.2. Participant Feedback Report

The participant feedback report can be generated in a similar manner to the lab results feedback report. An example of a participant feedback report is attached in Appendix 4.2.2.

1.9. CDART Training and Certification

Central CDART training took place in May 2012 via conference call. Two members of each site are required to be present. Those attending may provide additional training to other staff members at their sites. Follow-up conference call training sessions will be scheduled as needed. Monitoring site visits by CC personnel are scheduled to take place throughout the study; some CDART training can occur during these visits as well (depending on the circumstances and the perceived need for such training.)

1.10. Official Study Documents

Current versions of all study documents, protocol, data collection forms, MOP, user's guides, and other important documents are available on the study website at http://www.cscc.unc.edu/hchs. They are in the Study Members' area under the "Protocols and Manuals", and "Training Materials" pages. To access them the user must supply a username and password. Each document exists as an MS Word file and/or a PDF file. It is recommended that the PDF files be printed at the sites because formatting and special characters are retained. MS Word files are kept on the web site to facilitate working drafts as needed.

IMPORTANT: Versions of these documents that are designated as usable in the field (otherwise called "final") were sent to each study site as hard copy in an official SOL Youth Study Documents Notebook. One notebook will be provided to each Project Coordinator and Steering Committee Member. Section 13.8.1 describes the process of communicating updates to documents and of verifying receipt of communication on modifications.

1.10.1. Numbered Memos

The CC will routinely send emails or memos that are numbered and identified as "Numbered Memos". These memos are considered official documents and are to be stored at the back of the documents notebook. Updated information regarding the protocol, forms, MOP, QxQ's, and other documents being used in the field will be sent to the centers as Numbered Memos. Numbered Memos will be sent to all SOL Youth Project Coordinators and Steering Committee members. It is the site's responsibility to make sure this notification goes to all SOL Youth staff at each site that is affected. Each Project Coordinator must send email confirmation of receipt of a Numbered Memo to HCHSadminstration@unc.edu.

The numbered memo will instruct the recipient to print the updated version of the form, MOP, QxQ or other document from the Web and place this into the Site's Study Documents Notebook (Project Coordinator's notebook), replacing the older version. Numbered memos should be stored at the back of each binder form back to front with the most recent memo on top. Each site should provide archival storage of previous versions of documents according to their Institutional requirements. The CC will also keep all versions of official documents archived. Only memos that say "CC Memo #" should be filed in the Numbered Memos section.

The Project Coordinator's notebook (not the PI's notebook) is considered each site's official version of the documentation. The status of this notebook is monitored during any site monitoring visits.

1.10.2. Instructions for New / Corrected Materials

Forms: Any new or corrected forms will be available to print from the website. Forms should be replaced and copied for immediate use. Email confirmation must be sent to the CC (HCHSAdminstration@mail.cscc.unc.edu) when the revised forms are downloaded from the Internet.

Manual: The revised pages/chapters of the SOL Youth MOP 1 should be printed from the website and filed immediately in the MOP notebook. Email confirmation must be sent to the CC (HCHSAdminstration@mail.cscc.unc.edu) when the revised pages/chapters are downloaded.

QxQ's: Any new or corrected QXQ will be available to print from the website. They should be printed and filed immediately together with the appropriate form in the MOP binder. Email confirmation should be sent to the CC (HCHSAdminstration@mail.cscc.unc.edu) when the new QxQ's are downloaded.

1.10.3. Instructions for Outdated Materials

All outdated pages of the MOP, forms or QxQ's should be removed from the Documents Notebook as instructed in the Numbered Memo. Outdated materials should be archived according to each site's institutional requirements. All study materials are archived at the CC.

1.10.4. General Filing Instructions

All participants should have either a binder or file folder filed in alphabetical order by participant ID. If the center prefers to file by last name, there should be a cross-referenced list available with the corresponding ID number. It is important for centers to be able to communicate effectively with the CC by the participant's ID number. Data queries sent to the sites from the CC will only identify a participant by ID number. Remember, before sending any forms to the CC; blind (or mask) all personal information pertaining to the participant.

The safety and confidentiality of the study data and equipment is the responsibility of each study site. There is no need for participant data to be stored on portable memory devices such as flash drives or CDs. File cabinets with locks should be used for the storage of paper forms as well as the rooms containing these file cabinets. Computers containing participant information should also be locked.

2. Central Agency Data Management

Data management at each of the six reading centers will vary. This section describes the practices at each agency and the standards for data transfer established for SOL Youth.

2.1. Central Laboratory

The University of Minnesota Fairview Hospital Clinical Laboratory will receive, store and analyze the blood chemistries for the specimens transferred from the field centers. The central laboratory will send the results to the CC who will then send files for the feedback letters to the sites..

Each transfer from the central laboratory to the CC will comprise a single Excel spreadsheet. These worksheet names should not change without notice. Variables for each data set are defined in the first line of the sheet and are different for each set of data. The names in the spreadsheet are mapped to standard CDART variable names through a table called LABNAMES. It is important that once these names are agreed upon, they do not change without notice. The files are named according to study, lab and transfer sequence.

Central laboratory analyses are sent incrementally as samples are analyzed by the University of Minnesota Fairview Hospital Clinical Laboratory. If a record needs corrections to result values, it is included in a later transfer. The new record overwrites the original one.

The Excel files are uploaded to the CC using the CDART web site. Files are processed into the SOL Youth consolidated database overnight and available for reports the following day.

Clinic visit results are added to the clinical laboratory CLAA data table. The variable mapping is

HCHS-SOL Laboratory Tests, Reference Ranges, and Alert Values.

Test Names	Reference Ranges Males	Females	Age Ranges	Units
hs C-reactive protein	none/NR[a]	none/NR[a]	all	mg/L
e-Selectin	none/NR[a]	none/NR[a]	all	ng/mL
TNF-alpha	none/NR[a]	none/NR[a]	all	pg/mL
IL-6	none/NR[a]	none/NR[a]	all	pg/mL
Adiponectin	none/NR[a]	none/NR[a]	all	ng/mL
VWF (antigen)	none/NR[a]	none/NR[a]	all	%
PAI-1	none/NR[a]	none/NR[a]	all	U/mL
Insulin, fasting	none/NR[a]	none/NR[a]	all	pmol/L
Glucose, fasting[b]	60 – 99	60 – 99	all	mg/dL
Hemoglobin A1c	4.3 – 6.0	4.3 – 6.0	all	%
Total cholesterol[c]	<168	<177	5 – 9 years	mg/dL
	<173	<171	10 – 14 years	mg/dL
	<168	<176	15 – 19 years	mg/dL
Triglycerides[c]	<58	<74	5 – 9 years	mg/dL
	<74	<85	10 – 14 years	mg/dL
	<88	<85	15 – 19 years	mg/dL
LDL-cholesterol[c]	<103	<115	5 – 9 years	mg/dL
	<109	<110	10 – 14 years	mg/dL
	<109	<110	15 – 19 years	mg/dL
HDL-cholesterol[d]	>49	>48	5 – 9 years	mg/dL
	>46	>45	10 – 14 years	mg/dL
	>39	>43	15 – 19 years	mg/dL

[a] Non-reported test result so no reference range is necessary.

[b] Based on American Diabetes Association guidelines as reported in American Diabetes Association: Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2011;34 (Suppl. 1): S11–S61.

[c] Based on 75th percentile for children and adolescents as reported in Daniels SR, Greer, FR et al. Lipid Screening and cardiovascular health in childhood. Pediatrics 2008;122:198-208.

[d] Based on 25th percentile for children and adolescents as reported in Daniels SR, Greer, FR et al. Lipid Screening and cardiovascular health in childhood. Pediatrics 2008;122:198-208.

*These tests will be reported to the participants.

2.2. Nutrition and Supplements Reading Center

The University of Minnesota Nutrition Coordinating Center (NCC) supported by grants from NIH developed the NDS-R and DSAM software used to collect the 24 hour diet and supplement recall. The NDS software generates one file per participant interview. It is the responsibility of the local lead interviewer at each field center to review, over-read, and update the weekly collection of diet recall interviews. Once a week, a zip archive of the raw backup files will be uploaded to the CC using the CDART web site. The transfer files are named according to study and weekly date. For example a file created at the Bronx on October 22, 2012 would be named B_HCHS_20121022.

Files are later downloaded by NCC for nutrient coding. Once a month the NCC will transfer coded dietary interview data to the SOL Youth consolidated database.

3. SOL YOUTH COORDINATING CENTER DATA MANAGEMENT

Computers at the CC are connected via a Local Area Network. The network includes clustered database servers running the Novell network operating system and is connected to a Storage Area Network (SAN.) Clustered servers running the Microsoft network operating system provide web services for the data management system. The web servers are isolated by a router from the servers holding study data.

The consolidated database will be stored in a SQL-server database. Standard transaction validity checks will be applied to all updates to the database (e.g., to prevent the addition of records with duplicate keys, etc..) Audit logs from the CDART provide complete documentation for changes to the consolidated database. Backups of the consolidated database as well as processing reports are made daily. Once a month, the current backup tape is permanently archived at an off-site data storage facility. Periodically the consolidated database goes through a series of closure checks to ensure the completeness and correctness of data collection and processing. These checks are performed on a 'frozen' version of the database defined by a specific time cut point. Typical closure checks include classifying the universe of IDs, assuring all expected forms were received and assuring all queries generated were resolved.

3.1. Central Agency Data

Data from central agencies will be processed into the consolidated database every night or the night after receipt (if not entered via the CDART.) An automated process will read the data, perform key field and record level integrity checks and add valid records to the database. Data from each agency is processed and results reports are generated. The report lists key fields of records which were not processed due to error. It also lists total number of records processed without error. The processing reports are available to the agencies, the field centers and CC SOL Youth staff.

3.2. Field Center Data

Field center data, entered via the CDART, will be copied from the web server to the local area network (LAN) every night. From the files on the LAN, data will be retrieved into statistical analysis files for use in study reports for the steering committee and NHLBI.

3.3. Reports

The CDART will provide each field center with the ability to generate a variety of reports. These include participants contacted and examined, indicators of data quality, completion status of participant result reports and specimen tracking reports, among others. Such reports make it easy for study coordinators to monitor their center's performance and the timely identification and resolution of problems in data collection. The reports for study participants and their dentists or physicians are described in the CDART Manual under the "Reports" section. .

The field center data quality reports are complementary to the monthly Steering Committee reports. The former can be run in real time by field center staff and access up-to-date data stored in the consolidated data base. The Steering Committee reports will be produced monthly (or a schedule defined by the Steering Committee) and thus reflect the status of the study at the time of the most recent retrieval. Their purpose is to provide the Steering Committee and center investigators with performance information at all sites.

3.4. Data Security and Confidentiality

Data confidentiality and security will be applied at all levels of data acquisition, transfer and storage, and applied to all study agencies, from field centers to the CC. The CDART developed by the CC meets exacting data management standards of confidentiality, as well as HIPAA requirements. Beyond the password controlled access to the study equipment and the CDART, data collected at the field centers and in hospital record rooms are encrypted by the system and can only be decrypted for display on-screen by authorized study personnel. Personal identifiers are collected on separate forms (and transferred as separate, encrypted records.) The CC will be responsive to data confidentiality requirements originating from providers of medical care or IRBs, as needed to enable the work of the field centers. When paper data collection forms are used they will be retained at secure locations at the field centers until the Steering Committee acts on recommendations from the CC to dispose of such records (e.g., incremental data closure.) The secure storage and disposition of hard copy records at field centers will follow institutional procedures at each site.

The CDART server will be housed at the CC and will be exclusively managed by CC personnel. Measures to ensure the security of the data include: restricting access to users with valid IDs and passwords; using a firewall to restrict access to the web server and to shield the CC LAN from web users; using the secure sockets layer standard to provide encryption and user authentication. In accordance with CC standard operating procedures, system security logs and event logs are monitored daily to detect unauthorized attempts to access the system. The UNC Information Technology Systems group publishes a guide called "ITS Security at UNC Chapel Hill – Securing IIS". The CC follows these guidelines, which include closing unused ports; requiring user passwords to be long and difficult to guess; deleting certain files and subdirectories; and managing file and folder privileges.

All data transferred to the CC will be stored, processed, and analyzed within the CC office suite. At the CC, all access to office space containing data is controlled through locked doors. Visitors may enter only when accompanied by a CC escort. All office space remains locked after working hours. Access to computer data files is controlled by passwords released only to the CC personnel who use such files. In addition, data files with personal identifiers (and sensitive information per designation by a study's Steering Committee) are encrypted.

As standard practice, output mailed to a field center identifies participants only by ID number. No individually identifiable information will be distributed by the CC to any study agency other than the originating field center. Printed material containing confidential information is discarded through supervised loading, transportation, and storage using a chain of custody control process, until the material can be recycled into paper pulp.

It is a requirement for all CC staff to complete a confidentiality certification procedure upon employment. Policies regarding the confidential nature of the data collected, processed, and stored at the CC are explained to all personnel, who must then sign a "confidentiality certification" to be allowed access to confidential information. The CC reinforces the confidential nature of all study data at its staff meetings.

3.5. Data Retrieval and Statistical Computing

Data will be retrieved from the study database and converted into SAS files on a regular schedule (e.g., monthly.) The retrieved files will be stored as SAS datasets within a SAS data library. Most statistical computing will be done using SAS software. All statistical computing will be performed by a dedicated statistical programming staff, using a well-established statistical computing request system that has proven itself through use with many long-term, multi-center research projects managed by the CC. This system includes thorough documentation of requested computing, programming standards, naming conventions for datasets, programs and program results, inventorying and tracking of computing requests, procedures for program review, and permanent archival of completed programs, results, and datasets.

3.6. Database Closure

Data queries will be generated on a monthly basis, immediately following data retrieval. Typical data checks include classifying the universe of enrolled IDs, assuring all expected forms were received, performing consistency checks between related data fields, assuring all queries generated are resolved, etc. If there are unexpectedly high error rates for a site or a user, we explore the causes of the error and take corrective action, such as retraining personnel or making changes to the data management system. Research indicates that this comprehensive data checking, in combination with extensive real-time edits, can substitute for double data entry for data entered from paper forms.

Periodically the study's consolidated database is subjected to closure checks for completeness and accuracy of data collection and processing. These checks are performed on a "frozen" version of the database defined by a specific time cut point, and precede the use of data for publication. Typical closure checks include classifying the universe of IDs, assuring that all

expected forms were received and all queries were resolved, examining the consistency of items across forms and visits, and checking distributions of key variables for possible errors. Current plans entail closing the database in waves, one per examination year so that investigators will have access to interim results for study monitoring, review, and publication.

4. APPENDIX

4.1. Actical Export

Sending Actical (activity monitor) data to the CSCC is a two step process.

1) Run the **Actical Export** program to create a zip file of all files downloaded onto the computer since the last export.
2) Run the DMS Utility **Actical Upload** to send the zip file to the CC.

Export

After downloading the Actical monitors to your computer (location: C:\HCHS\Programfiles\Actical\), click on the icon on the Windows Desktop called 'Actical Export'.

If there are files to send, you will see the following screen:

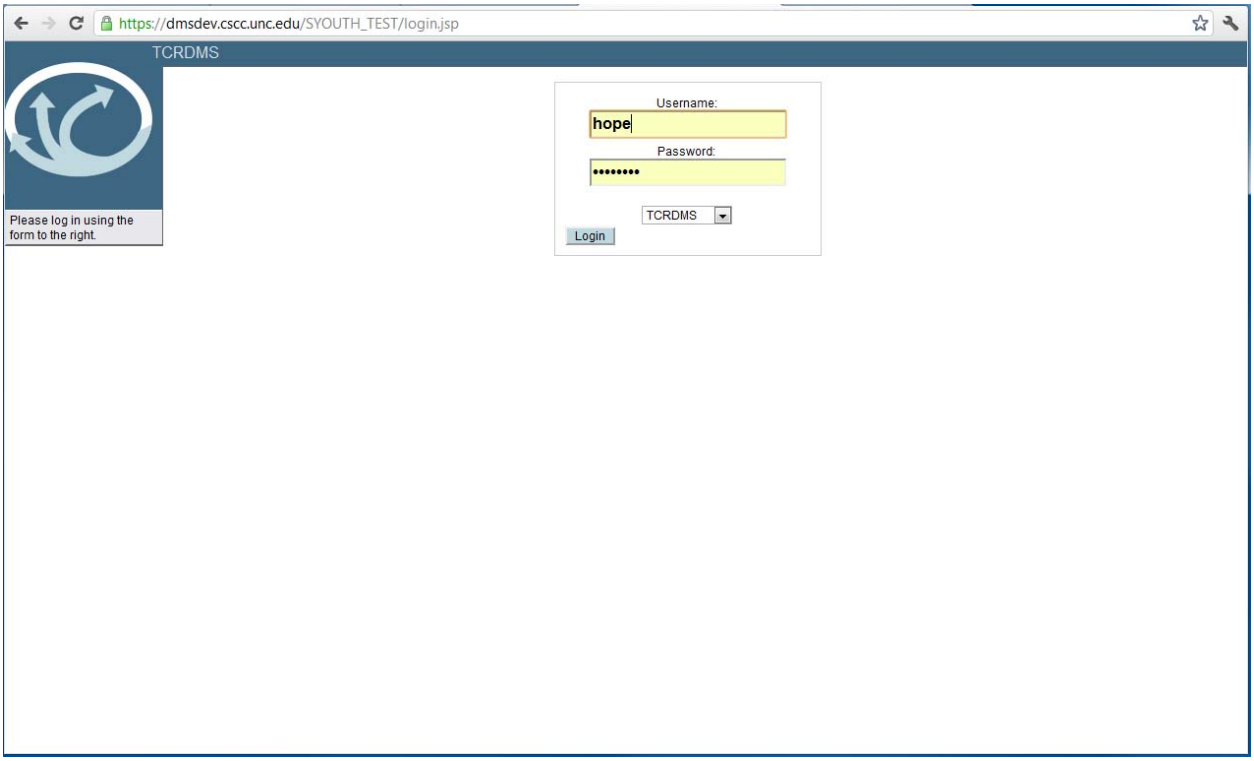
This screen tells you the name of the .zip file that was created – AB010006 – in the above example. The file name means this is Actical export (AB010006) for the Bronx (AB010006) computer number 1 (AB010006) export number 6 (AB010006.) This is the file to upload to the CC.

If there are no files to send, you will get the message:

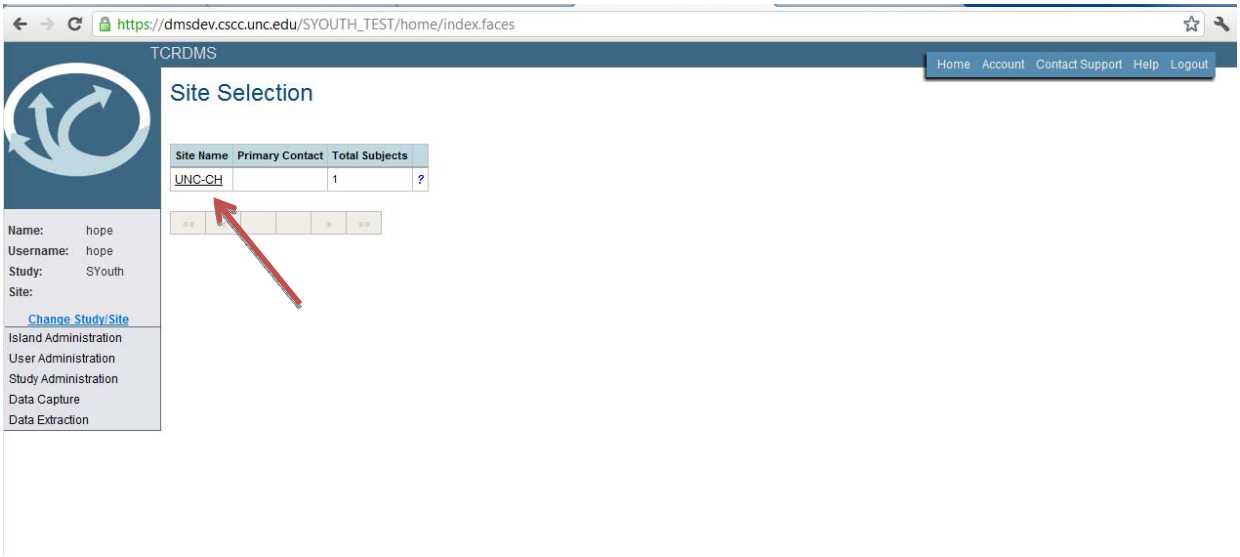
There are no files to export.

SOL-Youth Actical Data Upload

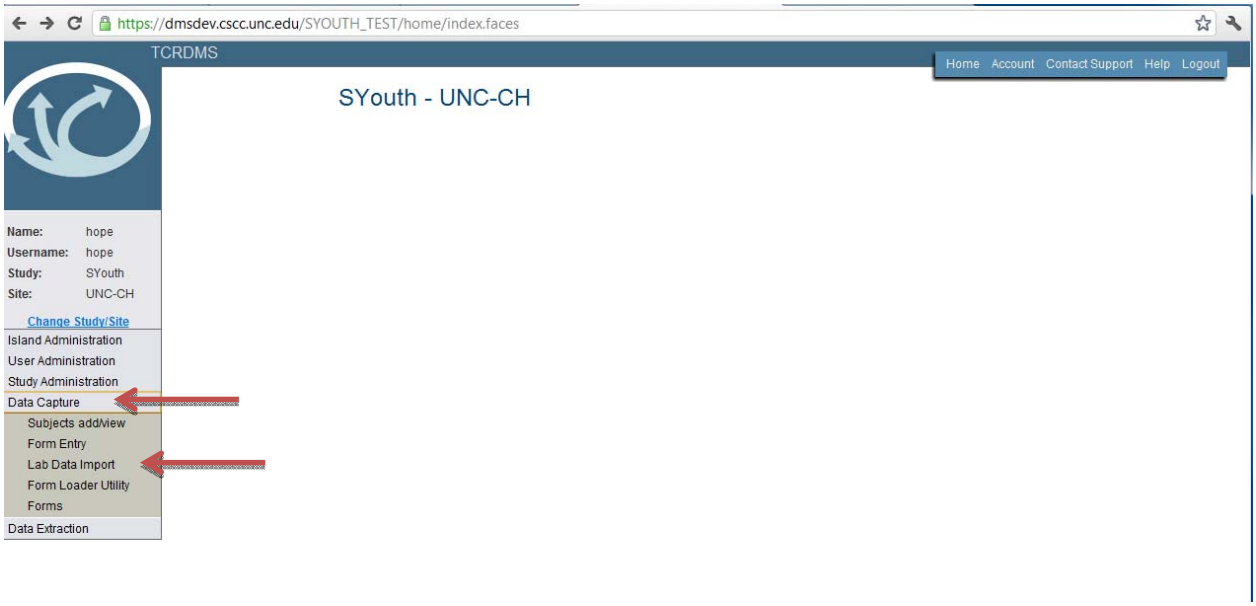
1) In a web browser (Chrome and Firefox work best) go to URL for study.
2) Login.



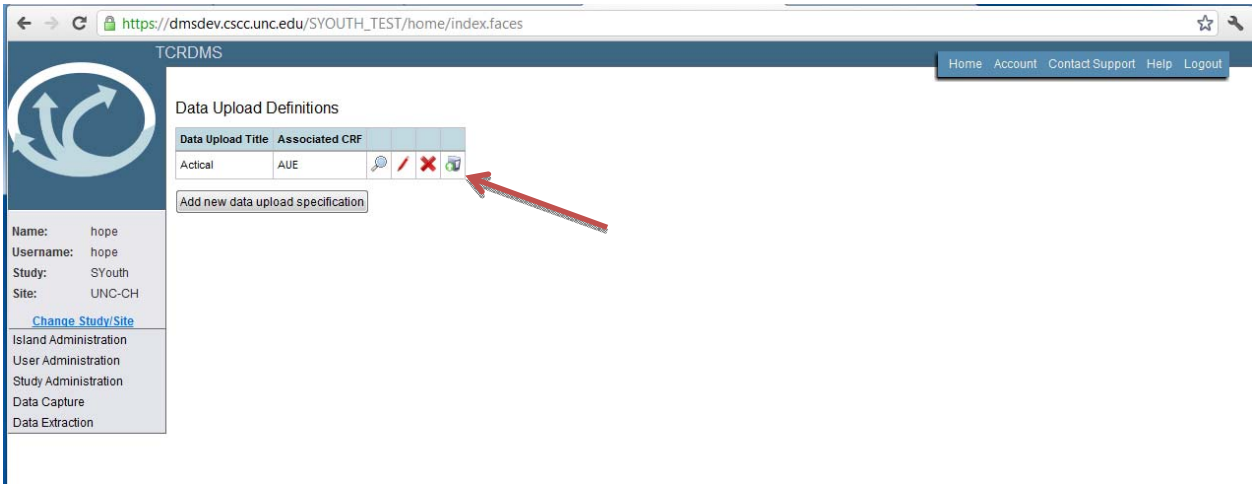
3) Select your site by clicking on the site name. (Here there is only one site because it is a test system.)



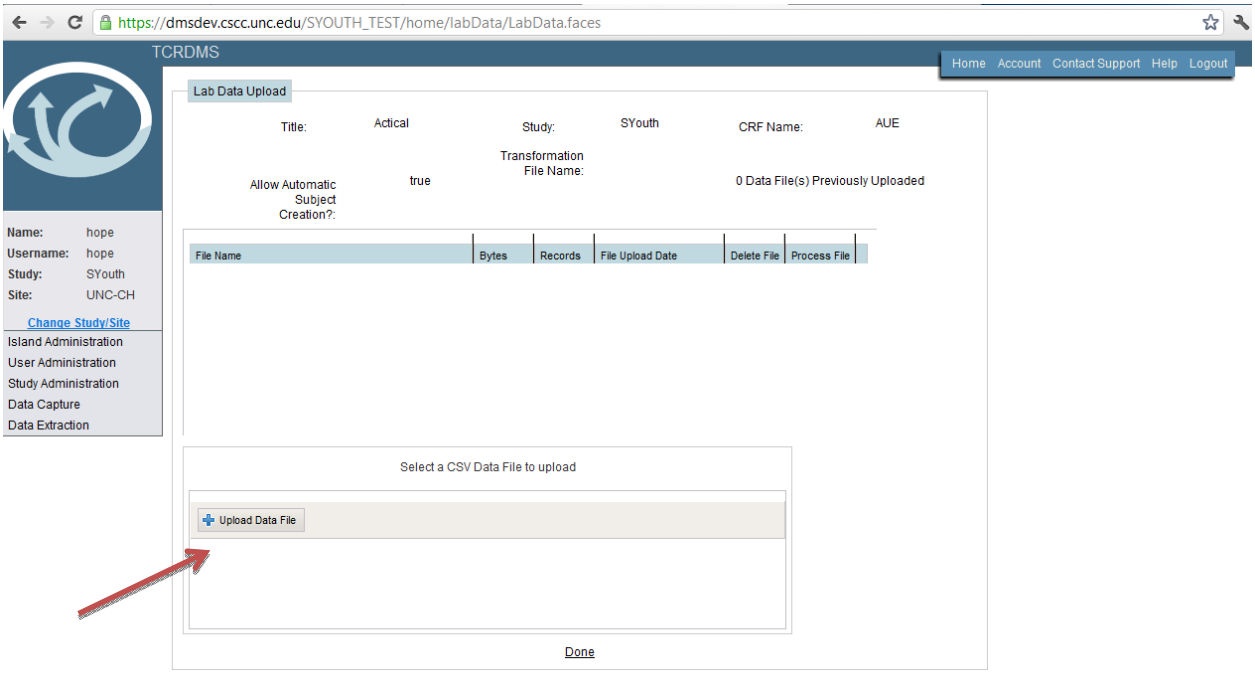
4) From left hand menu, click on ‘Data Capture’ to expand menu option.
5) Choose ‘Lab Data Import’.



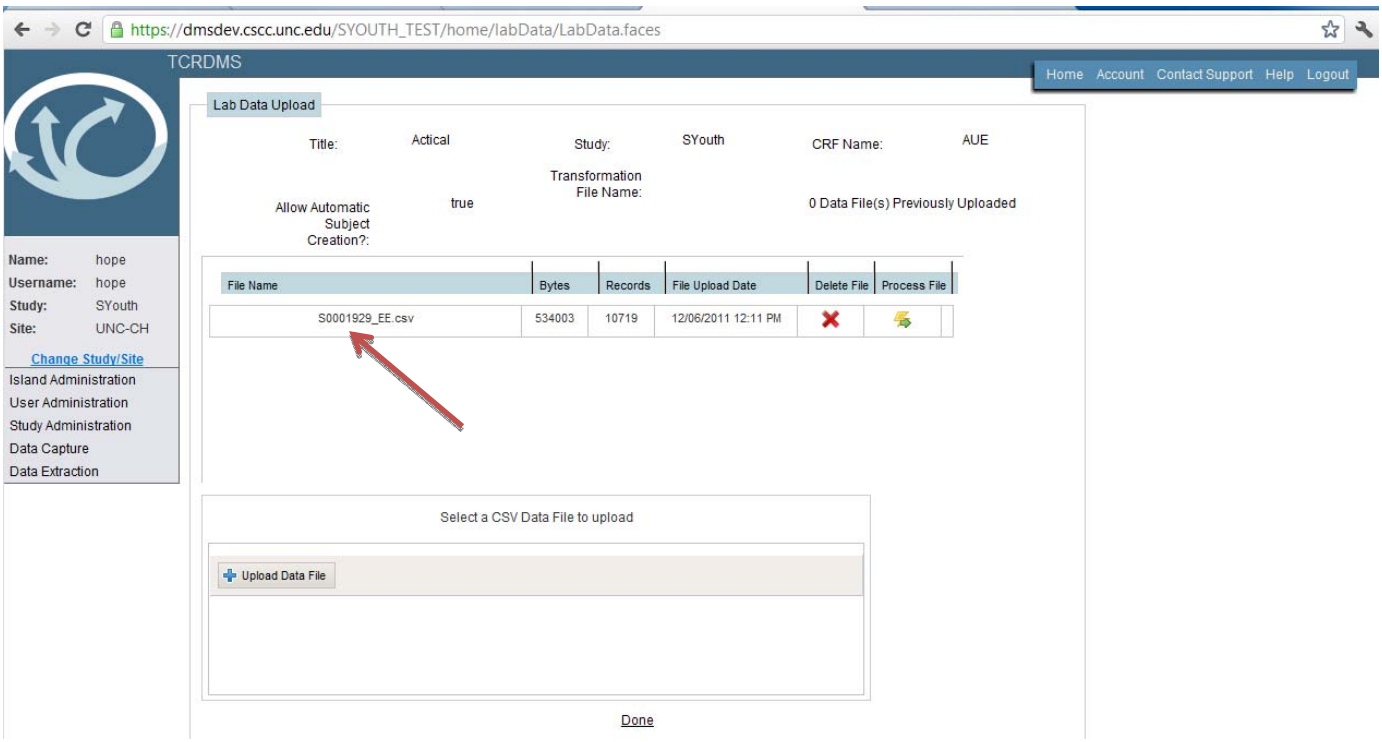
6) On the ‘Actical’ row, select the far right icon to upload files.



7) In the bottom panel, click ‘Upload Data File’.



8) From Windows Explorer, select the zipped Actical file you want to upload (location: C:\HCHS\Export\send\.)
9) Once the file is uploaded, it will appear in the upper panel. In the screen below, the file S0001929_EE.csv shows.



10) You can upload as many files as you want. Click ‘Done’ when finished.

4.2. Examples of Feedback Reports

4.2.1. Lab Results Feedback Report

SOL Youth Study

ID NUMBER: X0000000 STUDY YEAR: 01 SEQ# 01 Report Version Date: 1/19/12

SUMMARY OF SOL YOUTH STUDY LAB RESULTS FOR PARENTS/CAREGIVERS OF CHILD PARTICIPANTS

Participant's Name: _____ Age: XX Gender: X

Date of visit to the SOL Youth center: XX/XX/XXXX

Date/Time of Blood Collection: XX/XX/XXXX, XX:XX am

Date of Lab Report Release: XX/XX/XXXX Requester's Name: FC PI Name goes here

Testing Laboratory: University of MN Advanced Research and Diagnostic Laboratory
420 Delaware St. SE
Room L275 Mayo Bldg.
Minneapolis, MN 55455
Ph. (612) 273-3645

Dear Parent/Caregiver of _____:
Thank you very much for your and your child's participation in the SOL Youth Study. This report contains a summary of evaluations conducted during your child's recent visit to the SOL Youth center. This report contains results from your child's blood tests, including cholesterol, high density lipid (HDL), low density lipid (LDL), triglycerides, glycosylated hemoglobin A1c, and fasting blood sugar.

Laboratory Test Results:

Blood Tests	Reference Range	Test Result	Brief Interpretation
Total Cholesterol (mg/dL)	Age/Gender dependent so values will be auto-pop. from DMS	XXX	Auto-populated from DMS

ID NUMBER: X0000000
Participant's Name: _____ Age: XX Gender: X

Date of visit to the SOL Youth center: XX/XX/XXXX

Laboratory Test Results (cont.):

Blood Tests	Reference Range	Test Result	Brief Interpretation
LDL-Cholesterol (mg/dL)	Age/Gender dependent so values will be auto-pop. from DMS	XXX	Auto-populated from DMS
HDL-Cholesterol (mg/dL)	Age/Gender dependent so values will be auto-pop. from DMS	XXX	Auto-populated from DMS
Triglycerides (mg/dL)	Age/Gender dependent so values will be auto-pop. from DMS	XXX	Auto-populated from DMS
Glycosylated hemoglobin A1c (%)	4.3 – 5.6	XXX	Auto-populated from DMS
Fasting blood sugar (mg/dL)	60 - 99	XXX	Auto-populated from DMS

We encourage you to share this report with your child's health care provider. Again, thank you very much for your and your child's participation in the SOL Youth study. Do not hesitate to contact us at _____ at your local SOL Youth center if you have any questions. We look forward to your continued participation in the Hispanic Community Children's Health/Study of Latino Youth (SOL Youth) and related projects in the future.

Sincerely,

FC PI Name goes here (Auto-populated from DMS)

BRIEF EXPLANATIONS OF THE TESTS WE PROVIDE

Total cholesterol, LDL-cholesterol and triglycerides are the major fats in your blood. For more information please see http://www.americanheart.org/presenter.jhtml?identifier=1516

High density lipoprotein (HDL) cholesterol is also a blood fat; it appears to protect against hardening of the arteries. You can learn more about total cholesterol, LDL-cholesterol, HDL-cholesterol and triglycerides, at http://www.americanheart.org/presenter.jhtml?identifier=180

Glucose is a type of blood sugar that is the body's main source of energy and the amount in the blood is elevated in conditions such as diabetes. You can find more information about glucose levels and diabetes at http://www.diabetes.org/about-diabetes.jsp

If glucose levels are usually high in the blood this can be measured with glycosylated hemoglobin. A buildup of hemoglobin that is glycosylated usually indicates poorly controlled blood sugar levels in diabetics or undetected diabetes. More information about this test can be found at http://en.wikipedia.org/wiki/Glycosylated_hemoglobin

4.2.2. Participant Feedback Report

SOL Youth Study

ID NUMBER: X0000000 STUDY YEAR: 01 SEQ# 01 Report Version Date: 1/19/12

SUMMARY OF SOL YOUTH STUDY EXAM RESULTS FOR PARENTS/CAREGIVERS OF CHILD PARTICIPANTS

Participant's Name: _____ Age: XX Gender: X

Date of visit to the SOL Youth center: XX/XX/XXXX

Dear Parent/Caregiver of _____:

Thank you very much for your and your child's participation in the SOL Youth Study. This report contains a summary of evaluations conducted during your child's recent visit to the SOL Youth center. This report contains information on your child's weight and height, body mass index, and blood pressure.

Height, Weight, and Body Mass Index Results:

Weight: XXX kg (XXX lbs) Height: XXX cm (X ft, XX in)

Body Mass Index: XX.X BMI Percentile: XX

Body mass index (BMI) percentile is used to measure whether a child has a healthy weight compared with other youth of the same age and gender. Your child's BMI is normal if it is between the 6th and 84th percentile. If your child's BMI is below the 5th percentile, it is considered underweight. If your child's BMI is between the 85th and 94th percentile, it is considered overweight. If your child's BMI is above the 95th percentile, it is considered obese. If your child is underweight, overweight, or obese, you may want to discuss your child's results with your child's health care provider to consider options.

Blood Pressure Results:

Blood Pressure: XXX/XXX mm Hg (This is the average of three measurements taken after five minutes of rest)

Blood Pressure Percentile: XX/ XX percentile

Blood pressure (BP) percentile is used to measure how high the BP is compared to other youth of the same age, gender, and height. Blood pressure measures how hard the heart has to work to pump blood through the body. Your child's BP is considered normal if the BP is below the 90^{th} percentile. If the BP is between the 90^{th} and 95^{th} percentile it may indicate pre-hypertension. If the BP is at or above the 95^{th} percentile this may indicate hypertension. If your child's BP is not defined as normal, you may want to discuss your child's results with your child's health care provider to consider options.

We encourage you to share this report with your child's health care provider. Again, thank you very much for your and your child's participation in the SOL Youth study. Do not hesitate to contact us at _____ at your local SOL Youth center if you have any questions. We look forward to your continued participation in the Hispanic Community Children's Health/Study of Latino Youth (SOL Youth) and related projects in the future.

Sincerely,

FC PI Name goes here (Auto-populated from DMS)

BRIEF EXPLANATIONS OF THE TESTS WE PROVIDE

Obesity: The American Academy of Pediatrics recommends that health care providers to assess a child's risk for obesity and suggest weight control strategies for children who are overweight or obese. Because of the health consequences of being obese as a child, SOL Youth will report BMI categories per the American Academy of Pediatrics. You can find more information at http://www2.aap.org/obesity/about.html

Hypertension in children and adolescents continues to be defined as systolic BP and/or diastolic BP that is, on repeated measurement, at or above the 95th percentile. BP between the 90th and 95th percentile in childhood has been designated "high normal." To be consistent with the Seventh Report of the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7), this level of BP will now be termed "prehypertensive." For more information, please see http://www.nhlbi.nih.gov/health/prof/heart/hbp/hbp_ped.pdf

4.2.3. Participant Depression Feedback Report

SOL Youth Study

ID NUMBER: X0000000 STUDY YEAR: 01 SEQ# 01 Report Version Date: 01/18/12

SUMMARY OF SOL YOUTH DEPRESSION RESULTS FOR PARENTS/CAREGIVERS OF CHILD PARTICIPANTS

Participant's Name: _____ Age: XX Gender: X

Date of visit to the SOL Youth center: XX/XX/XXXX

Child Depression Inventory – Short Form (CDI-S)

CDI-S Score: XX

As part of today's assessment, your child reported a high number of symptoms associated with depression. On this screening measure, he/she reported symptoms greater than the 95% percentile. While this does not mean your child is depressed, it does suggest he/she has some concerns that should be evaluated by a health professional. We therefore recommend that you let your child's health care provider know of this result, and consider having a more formal assessment conducted by a mental health professional.

BRIEF EXPLANATIONS OF THE TESTS WE PROVIDE

The CDI-S form consists of 10 items each rated 0, 1, or 2; the range of scores for the scale is 0-20. Recent published reports indicate raw score cut-points of > 6 for boys and > 8 for girls exceeds the 95th percentile. These cut-points are used to identify children possibly at risk of depression, for whom additional evaluation by a health care provider may be indicated.