Organization of UNC CFAR Clinical Core

Joseph Eron
Core Director
Principal Investigator: HIV Clinical Research Unit (ACTG, HVTN, CHAVI)

Prema Menezes
Former Core Investigator

Byrd Quinlivan
Investigator
UNC ID Clinic Director

David Wohl
Investigator

Oksana Zakharova
Project Manager

Sonia Napravnik
Epidemiologist

UNC HIV Service

Christine Sun
Research Assistant

UNC ID Clinic

Ali Fokar
Data Manager/Programmer

Sam Stinnette
Statistician

Davina Chen
Clinic Screener

Heather Prince
Study Coordinator
Overview

- UCHCC Data Overview
- Data Management Objectives
- Regulatory Requirements
- Planning & Implementation
- Electronic Data Management Systems
- Case Report Forms Design (Guidelines)
- Data Quality Control
- Data Structure
UCHCC Data Overview

UCHCC: UNC CFAR HIV Clinical Cohort Study

Existing Institutional Electronic Databases
Medical Record Abstractions
Specimens (e.g., PBMC, Plasma, Cell Pellets)
CSDBS: Comprehensive In-person Patient Interviews
PRO: PROMIS, Patient Reported Outcomes
Clinic: Financial assessments, SAMISS, etc
RCTs: Studies, Labs, Treatments, etc
External: Nucleotide sequences
State / Federal: SSDI and NDI mortality data, Census data via Census Block Groups, Medicaid and Medicare data

Clinic:
- PSR: Patient Summary Report,
- PROs,
- Ryan White, CQI
Requests:
- Study feasibility,
- Grant submission,
- Enrollment,
- Data and specimen collection and provision
Research:
- Hypothesis generation,
- Investigator initiated,
- Graduate students and post-graduate fellows
Research Collaborations:
- National and International
Data Management Objectives

• The primary objective of Clinical Data Management (CDM) is to ensure timely delivery of high-quality data which are necessary to satisfy both good clinical practice (GCP) requirements and the statistical analysis and reporting requirements.

• The quality of the data validation process has a direct impact on the quality of research study.
IRB and HIPAA Requirements

- Researchers should prepare and submit their research protocols for IRB review and submit their HIPAA-related documents to the IRB at the same time.
  - Collect written authorization from patients for the release of their PHI.
  - IRB waiver from the authorization (Use of de-identified data)
- PHI that has been de-identified (stripped of a long list of identifiers) is not governed by HIPAA regulations.
- 2 cases under which IRB approval is not required but researcher must make representations under HIPAA if they are doing work with PHI.
  - Research on decedents.
  - Data review, preparatory to designing a research protocol.

http://www.unc.edu/hipaa/researchers.htm
Identifiers

- Names
- Geographic subdivisions smaller than a state
- Zip codes
- All elements of dates (DOB, DOD..)
- Telephone and Fax numbers
- Electronic mail addresses
- Social security numbers (SSN)
- Medical record numbers (MRN)
- Health plan beneficiary identifiers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- Device identifiers and serial numbers
- Web universal resource locators (URL)
- Internet protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images
- Any other number, characteristic or code that could be used by the researcher to identify the individual

http://research.unc.edu/files/2012/11/ccm3_0189871.pdf
Regulatory Requirements and Documentation

• U.S. Code of Federal Regulations (FDA regulated)
  • 21 CFR 11 Electronic Records (Electronic Data Capture and submission).

• International Conference on Harmonization (ICH) Good Clinical Practice (GCP) Guidelines
  • GCP 4.9 Records and Reports---(Investigators Responsibility)
ICH guidelines state: “The investigator should ensure the accuracy, completeness, legibility, and timeliness for the data reported to the sponsor in the CRFs and in all required reports.”

The investigator should ensure that any data reported on the CRF are consistent with the patient’s medical records and, where applicable, discrepancies should be explained.

NIH Requirements

• As of October 1, 2003, a data Sharing Plan is required to be included to all NIH grant applications $500,000 or more of funding.

• “In NIH's view, all data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data”

## Regulatory Requirements

**NC/UNC**

- **Sensitive (Y/N)**: Y Y Y Y Y
- **Applicable Laws and Regulations**:
  - PII: NC Identity Theft Protection Act
  - PHI: Health Insurance Portability and Accountability Act of 1996 (HIPAA)
  - Employee Data: GLBA State Personnel Act
  - FERPA: Family Educational Rights and Privacy Act (FERPA)
  - Non-public Information
- **Requires Encryption (Y/N)**: Y Y N N N
- **Has Applicable Security Standards (Y/N)**: Y Y Y Y Y
- **SAI Applicable for Servers (Y/N)**

Encrypt +password protect your Data: Portable devices, USB keys, Email

---

its.unc.edu/files/2012/03/Sensitive-Data-Examples
Transmitting research data via email

- Remove PHI and send de-identified data
- UNC Encrypted Email provides message encryption between the sender and recipient, and can be received by any email user
- This will encrypt the entire email including attachments.
- The subject trigger is: (secure).
- Also Password protect your data and send the passcode on a separate email

Source: http://help.unc.edu/help/unc-encrypted-email/
I need the p-value for the difference between drug A and drug B.
Case Report Forms

- Consistent look and feel
- Standard headers, footers, page numbering
- Instruction box for each form
- Standard format for branches (Skip Logic)
- Collect data outlined in the protocol
- Be clear and concise with your data questions
- Avoid duplication
- Request minimal free text responses

- Provide units to ensure comparable values
- Provide “choices” for each question
- Allow for Special codes
  - “d” Don’t know
  - “m” Missing data
  - “n” Not applicable
  - “r” Refused
  - “s” Too sick to respond
  - “?” Data item under query – requires follow-up
Types of questions

• Key fields – used to identify a unique record
• Multiple choice
  • Choose all that apply (Ex: Race)
  • Select only 1 (Ethnicity)
• Fill ins
  • Numeric and Character
  • specify # of characters
  • Specify decimal place
• Dates
• Open Ended
**Demographic Information Form (Di)**
Revision 1

The DRINK Study
Demographic Information Form

<table>
<thead>
<tr>
<th>Study ID</th>
<th>DOB</th>
<th>Visit ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Instructions:** Please fill out this form after determining that the patient is eligible for the study. Ask only one caregiver to respond to these questions. Read: "I am going to start by asking you some basic questions about your child, family and overall environment."

**Interviewer Guide:** Please try and use actual child’s and caregiver’s name as much as possible.

1. What is your relationship to the child?
   1) __Mother
   2) __Father
   3) __Other (please specify): ____________________

Please list everyone who lives in the home with the child by their relationship to the child and their age and sex, including yourself.

<table>
<thead>
<tr>
<th>A: Relationship code</th>
<th>B: If Other:Specify</th>
<th>C: Age (years)</th>
<th>D: Sex: 1=male, 2=female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Relationship codes: 1 = mother; 2 = father; 3 = sibling; 4 = grandparent; 5 = aunt/uncle; 6 = cousin; 7 = other

V:ShannonForms\Final Word docs\Di1.0.doc
Page 1 of 5

Created: 06/20/2006
Modified: 06/12/2006
DATA LIFE CYCLE

1. Protocol design
2. CRF Design
3. Metadata design
4. Database design
5. Data entry
6. Data collection
7. Database validation
8. Data cleaning
9. Dataset locked
10. Statistical analysis
11. Report writing
Electronic DM Systems

- Reduction in cycle time from protocol development to Statistical Analysis
- High Data Quality
- Lower Cost
- Improved Regulatory Compliance (complete audit trails)
- Facilitated clinical research monitoring capabilities

Options

- Commercial software packages Vs. in-house development
- Improved data integrity and quality, tracking techniques
# EDC Vs. CADE

## EDC

- Lab results from Webcis
- **Strengths:**
  - Eliminates data entry step
  - Timeliness
  - Accuracy
- **Weaknesses:**
  - Requires specialized computer programming expertise
  - Requires standards for representing clinical data (HL-7)
  - Requires willingness of systems managers at source of data to allow data connections

## CADE

- Chart abstraction form and CADET
- **Improve accuracy by:**
  - Double entry and file comparison (‘gold standard’ but inefficient and expensive)
  - Special technologies for referential integrity items (e.g., barcode visit and participant ID)
  - Data auditing and source document verification of scientifically important variables
EDC: Participant entry

- Example – PRO collection
- use thin tablets
- Strengths
  - If well designed, eliminates data entry step
  - Can add multimedia explanations and tutorials
  - Can be more enjoyable for study participants than paper forms
- Weaknesses
  - Requires basic computer skills
  - Requires literacy skills
  - Requires staff assistance and verification
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>H69</td>
<td>1.037</td>
<td>82.944</td>
<td>1.074</td>
<td>59.321</td>
<td>0.945</td>
<td>81.108</td>
<td>1.372</td>
<td>92.836</td>
</tr>
<tr>
<td>5</td>
<td>H82</td>
<td>0.688</td>
<td>63.338</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>H128</td>
<td>0.72</td>
<td>68.165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H146</td>
<td>0.901</td>
<td>78.758</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>H187</td>
<td>0.836</td>
<td>74.275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>H209</td>
<td>0.716</td>
<td>67.949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>H345</td>
<td>0.807</td>
<td>72.237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>H446</td>
<td>0.912</td>
<td>76.145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>H526</td>
<td>0.883</td>
<td>74.674</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H774</td>
<td>0.892</td>
<td>79.231</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>H847</td>
<td>0.38</td>
<td>48.294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>H1092</td>
<td>0.692</td>
<td>70.134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>H1618</td>
<td>0.821</td>
<td>77.187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>H1930</td>
<td>0.779</td>
<td>76.673</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LX22</td>
<td>0.211</td>
<td>33.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LX33</td>
<td>0.238</td>
<td>35.864</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LX36</td>
<td>0.659</td>
<td>68.484</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Electronic DMS

<table>
<thead>
<tr>
<th>Feature</th>
<th>Teleform</th>
<th>MS Access</th>
<th>REDCap</th>
<th>CDART</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires Data Entry</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>web-based.</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data Encryption</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Audit trails</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>User Management</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multi-site access and data entry</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Auto-validation, branching logic, and stop actions.</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mid-study modifications. You may modify the database or survey at any time during the study.</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data comparison functions/Double data entry</td>
<td>✗</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Real-time Data reports</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Notification System</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Scalability</td>
<td>✗</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Custom Functionalities</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
TeleForms

Read as 02593-6-0 instead of 02693-6-0

Read as 02103-6-0 instead of 02103-5-0
DMS Validation

• Enter dummy data provided by the developer.
• Test all navigation buttons.
• Ensure that data entered through the data entry screen are saved appropriately and can be browsed, changed, and/or deleted.
• Validate any data integrity constraints or checking routines that execute during data entry.
• Document all testing performed as part of the validation process.
• The first week that the data system is used in the field should be considered a Beta-test period. During this period, end-users in the field should enter fictitious data to test the system in the environment that it will actually be used.
DATA LIFE CYCLE

Protocol design → CRF Design → Metadata design → Database design

Data entry → Data collection → Database validation

Data cleaning → Dataset locked → Statistical analysis → Report writing
Data entry following paper CRF collection

- Example - chart abstraction form and CADET
- Average keystroke error rates will be 0.1% to 1%, depending upon data type
  - Improve accuracy over baseline by:
    - Double entry and file comparison ('gold standard' but inefficient and expensive)
    - Special technologies for referential integrity items (e.g., barcode visit and participant ID)
    - Event-driven auditing and source document verification of scientifically important variables
Mapping DB Names to Form Items

- **One-to-one** mapping between:
  - DB data item names
  - Item numbers on the data forms and revisions
- Required by analyst to make sense of the data
- If variable naming conventions are followed and transparent, then a collection of all the forms will also serve as the data dictionary
DATA LIFE CYCLE

Protocol design → CRF Design → Metadata design → Database design

Data entry ← Data collection ← Database validation

Data cleaning → Dataset locked → Statistical analysis → Report writing
Data Quality Control

• Raw data files are never altered
• Changes made via program (Eg: SAS, Stata)
• Check for . . .
  • Inconsistencies
  • Duplicate or missing IDs or records
  • Out-of-range errors
  • Logical errors
    • Within data table and between tables
• Check by writing computer code
  • Do not use eye-ball method
Data Structures

- Flat file
  - One record per person (or unit of analysis)
- Normalized
  - Divided into many tables via well defined relationships to reduce redundancy
  - Standard among database administrators
  - Relational databases
- Form based
Flat File

SANDSID
RIGHT_ARM_CIRC_V1
CUFF_SIZE_V1
RIGHT_ARM_V1
STANDING_TIME_V1
TWO_MIN_SYS_V1
TWO_MIN_DIAS_V1
ANTHROPOMETRICS_V1
STAFF_CODE_V1
VISIT_DATE_V1
SYS_1_V1
DIAS_1_V1
SYS_2_V1
DIAS_2_V1
SYS_3_V1
DIAS_3_V1
WEIGHT_V1
WAIST_V1

Continued...
RIGHT_ARM_CIRC_V2
CUFF_SIZE_V2
RIGHT_ARM_V2
STANDING_TIME_V2
TWO_MIN_SYS_V2
TWO_MIN_DIAS_V2
ANTHROPOMETRICS_V2
STAFF_CODE_V2
VISIT_DATE_V2
SYS_1_V2
DIAS_1_V2
SYS_2_V2
DIAS_2_V2
SYS_3_V2
DIAS_3_V2
WEIGHT_V2
WAIST_V2
...repeat for every visit
<table>
<thead>
<tr>
<th>ID</th>
<th>Test_Date</th>
<th>CD4</th>
<th>Medications</th>
<th>Med_StartDate</th>
<th>Med_endDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3/5/1997</td>
<td>131</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3/5/1997</td>
<td>131</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4/6/1997</td>
<td>134</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4/6/1997</td>
<td>134</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>11/5/1997</td>
<td>132</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>11/5/1997</td>
<td>132</td>
<td>Atazanavir (Reyataz)</td>
<td>1/20/2012</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>11/5/1997</td>
<td>132</td>
<td>Tenofovir (Viread)</td>
<td>9/5/2006</td>
<td>39663</td>
</tr>
<tr>
<td>32</td>
<td>11/5/1997</td>
<td>132</td>
<td>Saquinavir (Fortovase, Invirase)</td>
<td>39555</td>
<td>39663</td>
</tr>
<tr>
<td>33</td>
<td>11/5/1997</td>
<td>132</td>
<td>Ritonavir Boost (Norvir)</td>
<td>39555</td>
<td>39663</td>
</tr>
<tr>
<td>34</td>
<td>11/5/1997</td>
<td>132</td>
<td>Etivirite (DMP266, Sustiva)</td>
<td>39555</td>
<td>39663</td>
</tr>
<tr>
<td>35</td>
<td>11/5/1997</td>
<td>132</td>
<td>Atraviir (Glaxi)</td>
<td>39555</td>
<td>39663</td>
</tr>
<tr>
<td>36</td>
<td>11/5/1997</td>
<td>132</td>
<td>Truvir (AZT, 3TC, Ativanir)</td>
<td>39011</td>
<td>39065</td>
</tr>
</tbody>
</table>
A Day in the Life of Databert

BAM BANG! BOOM!

IS THAT QUERY RESULT BACK YET?

DATAlegro
(949) 330-7690
www.datalegro.com
SELECT  a.patient_key, a.regimen, a.medication, a.startDate, a.enddate, b.test_number, b.test_date, b.test_result
FROM    APPDATA.MEDICATIONS as a
         FULL Join
         WDATA.LABORATORY_FACT_TABLE as b
On      a.patient_key = b.patient_key
Normalized Datasets

MedHis
- SANDSID
- SCREENID
- HIBP
- HIBPAGE
- DIABETES
- ALLERGIES
- ARTHRITIS
- FRACTURES
- FRACTURE_LOCATION
- GALLSTONES
- PANCREATITIS
- CONSTIPATION
- CHRONIC_DIARRHEA
- CIRRHOSIS

MedHis_Diab
- SANDSID
- SCREENID
- AGE
- INSULIN
- ORAL
- DIET
- EXERCISE
- NOTHING
- OTHER

MedHis_Allergy
- SANDSID
- SCREENID
- SEASONAL
- ANYMEDS

MedHis_Allergy_Med
- SANDSID
- SCREENID
- MEDICATION
- DOSE
- START_DT
- END_DT

Exam
- SANDSID
- VISIT
- RIGHT_ARM_CIRC
- CUFF_SIZE
- RIGHT_ARM
- STANDING_TIME
- TWO_MIN_SYS
- TWO_MIN_DIAS
- ANTHROPOMETRICS
- STAFF_CODE
- VISIT_DATE

Exam_BP
- SANDSID
- VISIT
- MEASUREMENT_NUM
- SYS
- DIAS

Exam_Anthro
- SANDSID
- VISIT
- WEIGHT
- WAIST
Merging data sets

• Understand key fields
  • Combination of fields that define a unique record
  • Merge based on key fields

• Types of merges
  • Merging data sets with the same key fields
    • Inner join
    • Outer join (full join)
    • Right/Left join
  • Merging data sets with different key fields
    • One to many join
    • Many to many
Creating Analysis Database

• Freeze date
  • The date when the study databases are copied to a new location and no longer updated
  • Ability to obtain identical analysis at any time in the future
  • Need to create analysis data sets that are amalgamation of form revisions and edits
Collaborative Studies Coordinating Center:  [http://www2.cscc.unc.edu/home/](http://www2.cscc.unc.edu/home/)

Data Management Courses:

- **ODUM institute:**  [http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=667](http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=667)

UNC CFAR HIV/AIDS Clinical and Research Database:  [http://cfar.med.unc.edu/content/clinical-core](http://cfar.med.unc.edu/content/clinical-core)

[http://tracs.unc.edu/](http://tracs.unc.edu/)

[https://virtuallab.unc.edu/](https://virtuallab.unc.edu/)
Acknowledgment

- Sonia Napravnik
- Joe Eron
- Michael Hudgens
- Katie Mollan
- David Rosen

- Carol Golin
- Catherine Grodensky
- Oksana Zakharova
- Sam Stinnette
- Christine Sun