

Data Management in Clinical Research

UNC Center for AIDS Research

August 30, 2013

Ali Fokar

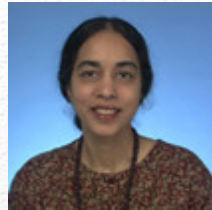


UNC
CFAR
Center for AIDS Research

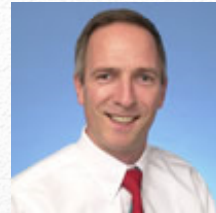
Organization of UNC CFAR Clinical Core



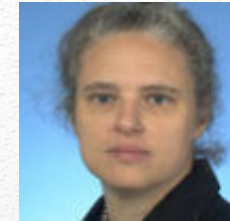
David Wohl
Investigator



Prema Menezes
Former Core
Co-Director



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



Byrd Quinlivan
Investigator
*UNC ID
Clinic
Director*



Sonia Napravnik
Epidemiologist



Oksana Zakharova
Project
Manager



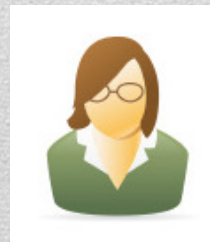
UNC ID Clinic



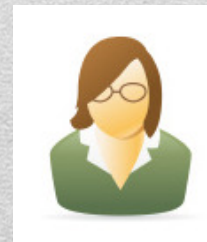
Ali Fokar
Data
Manager/
Programmer



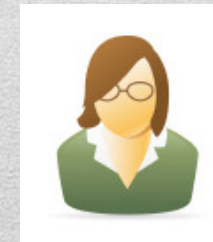
Sam Stinnette
Statistician



Davina Chen
Clinic Screener



Christine Sun
Research
Assistant



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

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

UCHCC: UNC CFAR HIV Clinical Cohort Study

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.

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

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)
-

International Conference on Harmonization GCP 4.9

- 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

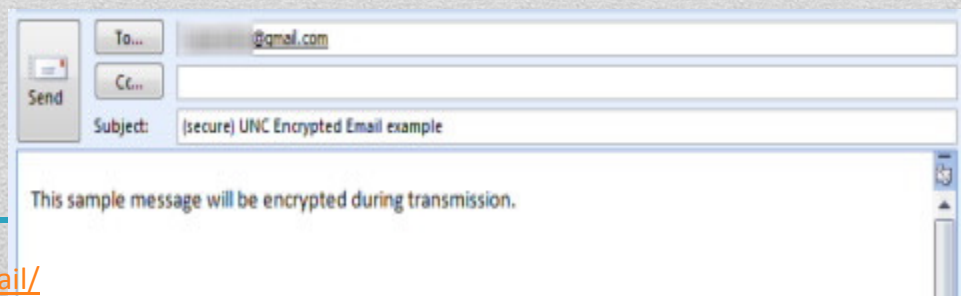
NC/UNC

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

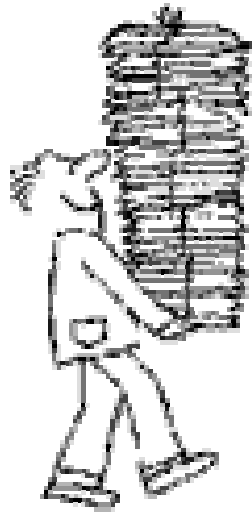
	PII	PHI	Employee Data	FERPA	Non-public Information
Sensitive (Y/N)	Y	Y	Y	Y	Y
Applicable Laws and Regulations	NC Identity Theft Protection Act Gramm Leach Bliley Act (GLBA)	Health Insurance Portability and Accountability Act of 1996 (HIPAA)	GLBA State Personnel Act	Family Educational Rights and Privacy Act (FERPA)	
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)¹	Y	Y	Y	Y	Y

Transmitting research data via email

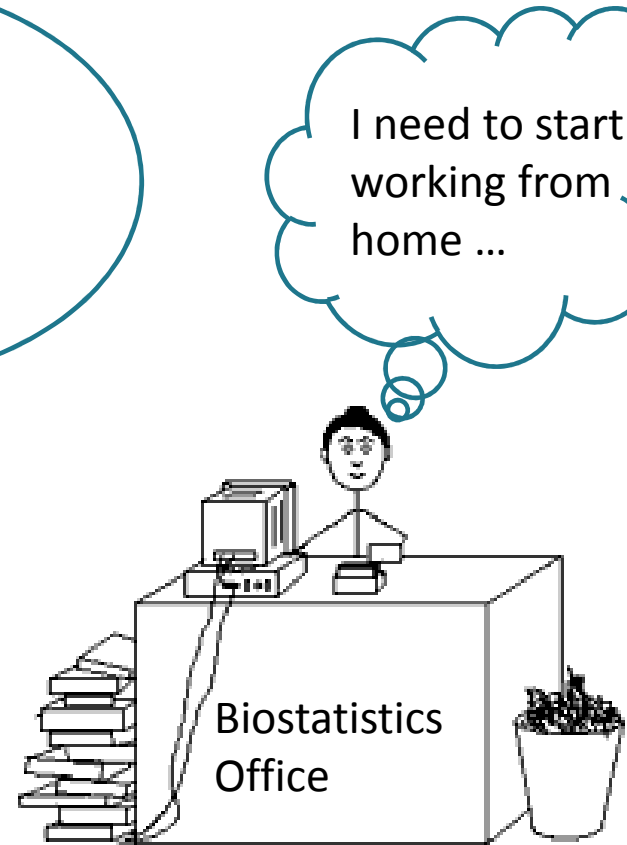
- Remove PHI and send de-identified data
- UNC Encrypted Email provides message encryption between the sender and recipient, and can be received by any email user
- This will encrypt the entire email including attachments.
- The subject trigger is: (secure).
- Also Password protect your data and send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



Source :<http://help.unc.edu/help/unc-encrypted-email/>

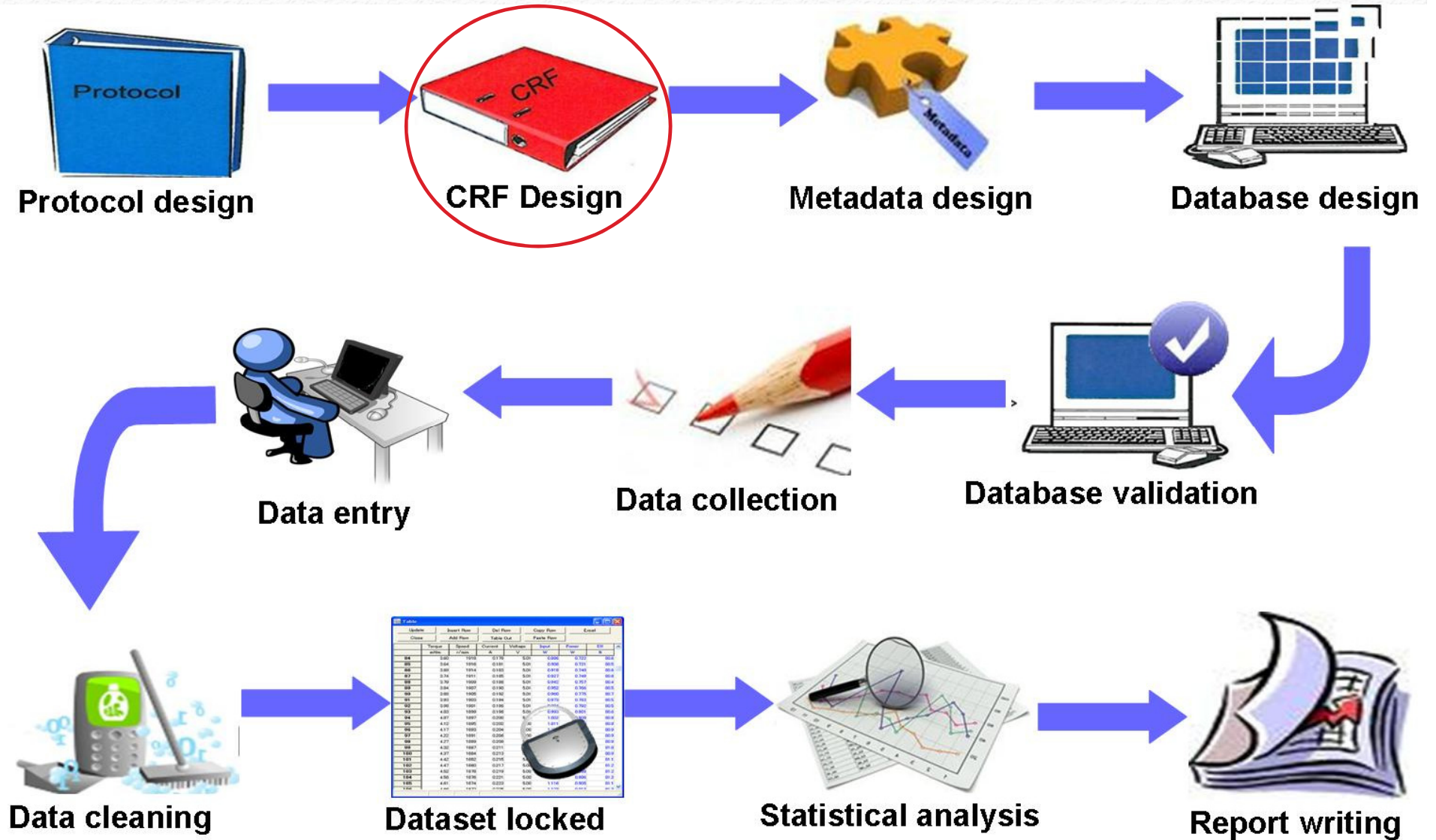


I need the
p-value for
the difference
between drug
A and drug B



I need to start
working from
home ...

DATA LIFE CYCLE



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 questions
 - 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
-



The DRINK Study
Demographic Information Form

Study ID _____ DOB / / dd mon Visit ID

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?

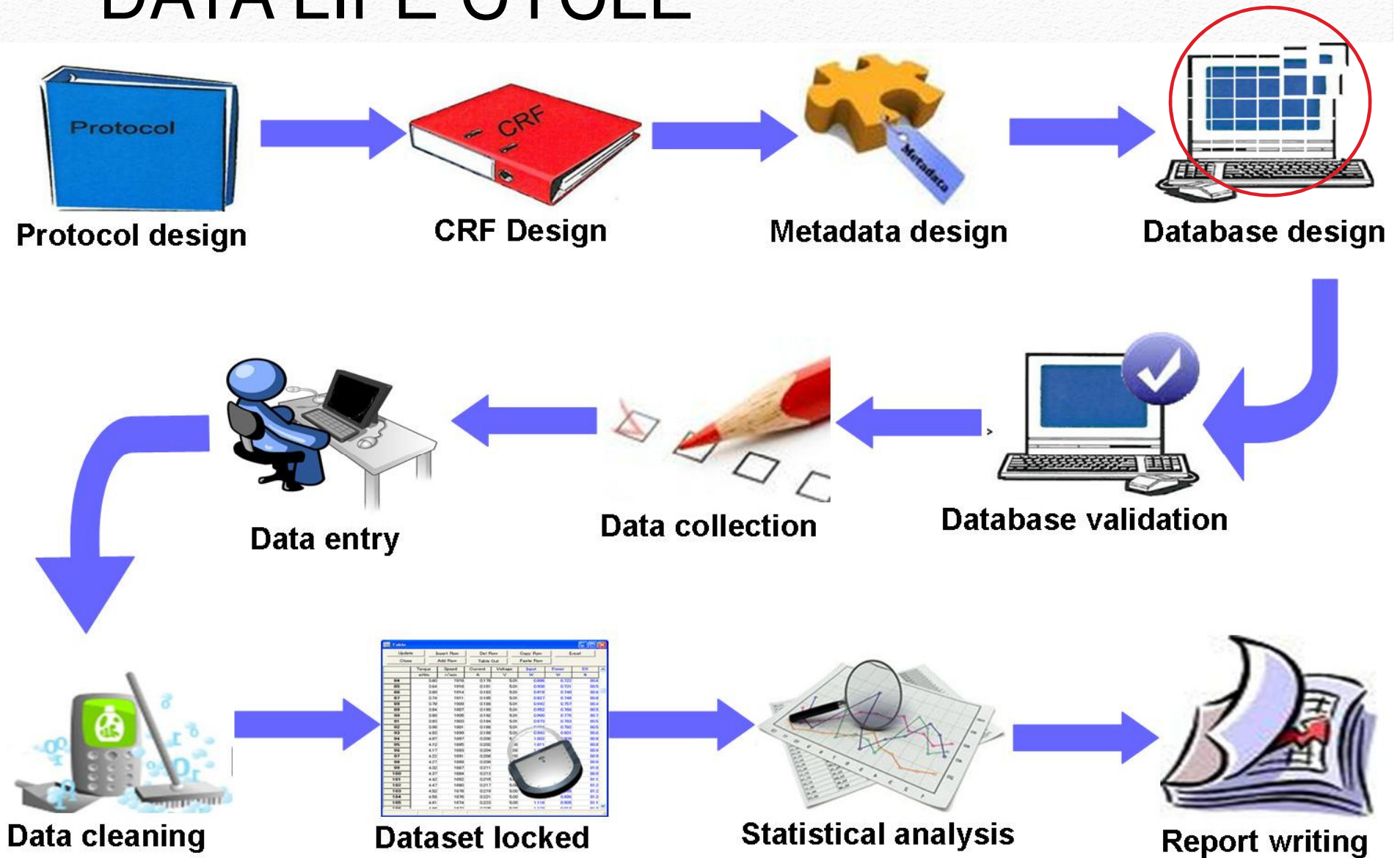
- 1) ___ Mother
- 2) ___ Father
- 3) ___ Other: (please specify): _____

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

	A* Relationship code	B If Other, Specify	C Age (years)	D Sex 1=male, 2=female
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				

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

DATA LIFE CYCLE



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

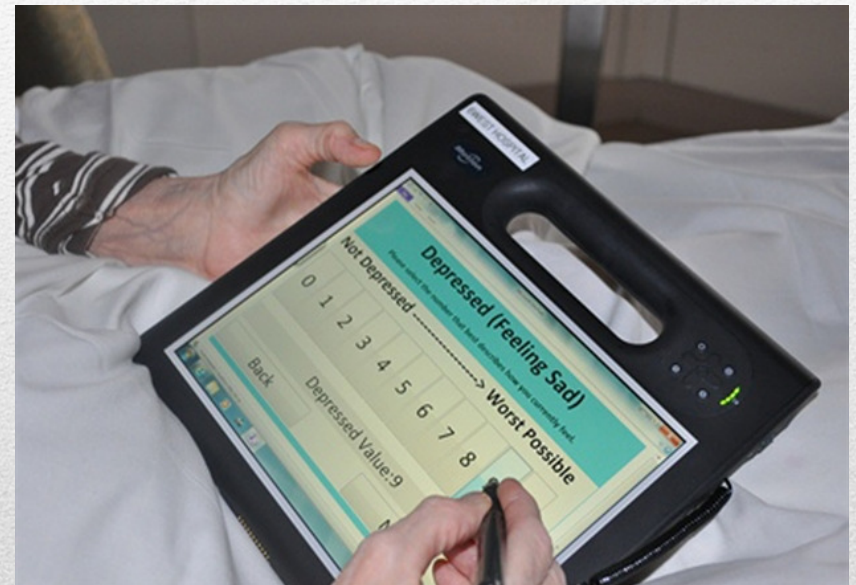
- Lab results from Webcis
- Strengths:
 - Eliminates data entry step
 - Timeliness
 - Accuracy
- Weaknesses:
 - Requires specialized computer programming expertise
 - Requires standards for representing clinical data (HL-7)
 - Requires willingness of systems managers at source of data to allow data connections

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



MS Excel

density_scans_compiled.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	lane	line	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.	OD	Rel. Intens.
4	1	H69	1.037	82.944	1.074	59.321	0.945	81.108	1.372	92.836	0.061	33.214	-	-			-	-
5	2	H82	0.688	63.338	1.1													
6	3	H128	0.72	68.165	1.1													
7	4	H146	0.901	78.758	1.0													
8	5	H187	0.836	74.275	1.1													
9	6	H209	0.716	67.949	1.0													
10	7	H345	0.807	72.237	1.2													
11	8	H446	0.912	76.145	0.9													
12	9	H526	0.883	74.674	0.9													
13	1	H774	0.892	79.231	0.7													
14	2	H847	0.38	48.294	0.7													
15	3	H1092	0.692	70.134	0.7													
16	4	H1618	0.821	77.187	0.8													
17	5	H1930	0.779	76.673	0.8													
18	6	LX22	0.211	33.416	1.0													
19	7	LX33	0.238	35.864	1.0													
20	8	LX36	0.659	68.484	0.9													
21	CF7; ignored)																	
22																		
23																		
24																		
25																		
26																		
27																		
28																		
29																		
30																		

density_scans_compiled.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Inference JMP

Clipboard Font Alignment Number Styles

Sort Warning

Microsoft Office Excel found data next to your selection. Since you have not selected this data, it will not be sorted.

What do you want to do?

Expand the selection

Continue with the current selection

Sort Cancel

Electronic DMS

	Teleform	MS Access	REDCap	CDART
Requires Data Entry	X	√	√	√
web-based.		X	√	√
Data Encryption		X	√	√
Audit trails		X	√	√
User Management		X	√	√
Multi-site access and data entry	√	√	√	√
Auto-validation, branching logic, and stop actions.	X	√	√	√
Mid-study modifications. You may modify the database or survey at any time during the study.	√	X	√	√
Data comparison functions/Double data entry		X	√	√
Real-time Data reports	X	√	√	√
Notification System	X	X	√	√
Scalability		X	√	√
Custom Functionalities	X	X	√	√

TeleForms

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

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

Medical History

Date of Visit: 09 - 06 - 2007
 Visit Code: 02 - 00 Staff Code: 251

Identification Number: 02593-6-0

PMH v. 2.0 Page 1 of 1

This form contains selected questions from PMH version 1.0
 Mafunso otsatirawa ndi okhudza za matenda omwe munali nawo kapena muli nawo. Ndiyamba ndi zinthu zomwe tikufuna kudziwa ngati munayamba mwakumanapo nazo.

1. Kodi pali mankhwala amene simuyanjana nawo?
 inde → **Ngati ndi choncho mankhwala ake ndi ati**

Kodi . . .

3. Munadwalapo chifuwa chachikulu? inde ayi sindikudziwa

4. Munadwalapo matenda achikasu? inde ayi sindikudziwa

5. Munadwalapo zilonda za mkamwa? inde ayi sindikudziwa

6. Munadwalapo mashingozi? inde ayi sindikudziwa

7. Munadwalapo matenda othamanga magazi? inde ayi sindikudziwa

8. Munadwalapo matenda a shuga? inde ayi sindikudziwa

9. Munadwalapo matenda a mtima? inde ayi sindikudziwa

10. Munadwalapo matenda a impsyo? inde ayi sindikudziwa

11. Munadwalapo matenda akubongo monga khunyuu? inde ayi sindikudziwa

12. Munadwalapo matenda amphumu? inde ayi sindikudziwa

13. Munadwalapo matenda wotha magazi pafupipafupi? inde ayi sindikudziwa

14. Munadwalapo matenda opatsirana pogonana pa miyezi khumi ndi iwiri yepitayi? inde ayi sindikudziwa

Kodi munalandirapo katemera wa . . .

15. Tetanus (TTV)? inde ayi sindikudziwa

16. BCG? inde ayi sindikudziwa

17. Any other vaccine? inde ayi sindikudziwa

334 Reviewer Code: 6890155745

Past Medical History

Date of Visit: 22 - 12 - 2006
 Visit Code: 02 - 00 Staff Code: 457

Identification Number: 02103-6-0

PMH v. 2.0 Page 1 of 1

This form contains selected questions from PMH version 1.0
 Mafunso otsatirawa ndi okhudza za matenda omwe munali nawo kapena muli nawo. Ndiyamba ndi zinthu zomwe tikufuna kudziwa ngati munayamba mwakumanapo nazo.

1. Kodi pali mankhwala amene simuyanjana nawo?
 inde → **Ngati ndi choncho mankhwala ake ndi ati**

Kodi . . .

3. Munadwalapo chifuwa chachikulu? inde ayi sindikudziwa

4. Munadwalapo matenda achikasu? inde ayi sindikudziwa

5. Munadwalapo zilonda za mkamwa? inde ayi sindikudziwa

6. Munadwalapo mashingozi? inde ayi sindikudziwa

7. Munadwalapo matenda othamanga magazi? inde ayi sindikudziwa

8. Munadwalapo matenda a shuga? inde ayi sindikudziwa

9. Munadwalapo matenda a mtima? inde ayi sindikudziwa

10. Munadwalapo matenda a impsyo? inde ayi sindikudziwa

11. Munadwalapo matenda akubongo monga khunyuu? inde ayi sindikudziwa

12. Munadwalapo matenda amphumu? inde ayi sindikudziwa

13. Munadwalapo matenda wotha magazi pafupipafupi? inde ayi sindikudziwa

14. Munadwalapo matenda opatsirana pogonana pa miyezi khumi ndi iwiri yepitayi? inde ayi sindikudziwa

Kodi munalandirapo katemera wa . . .

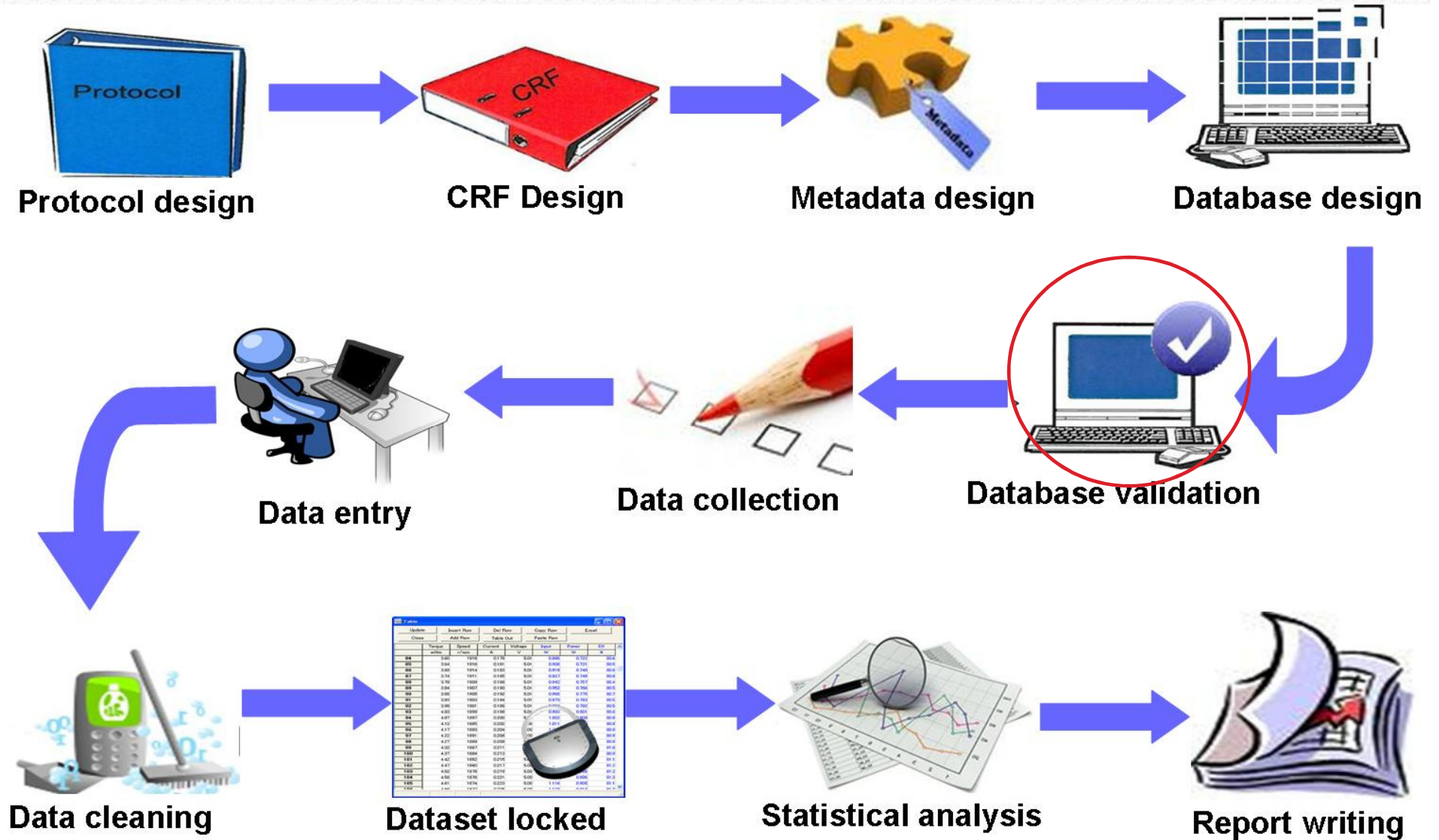
15. Tetanus (TTV)? inde ayi sindikudziwa

16. BCG? inde ayi sindikudziwa

17. Any other vaccine? inde ayi sindikudziwa

174 Reviewer Code: 6890155745

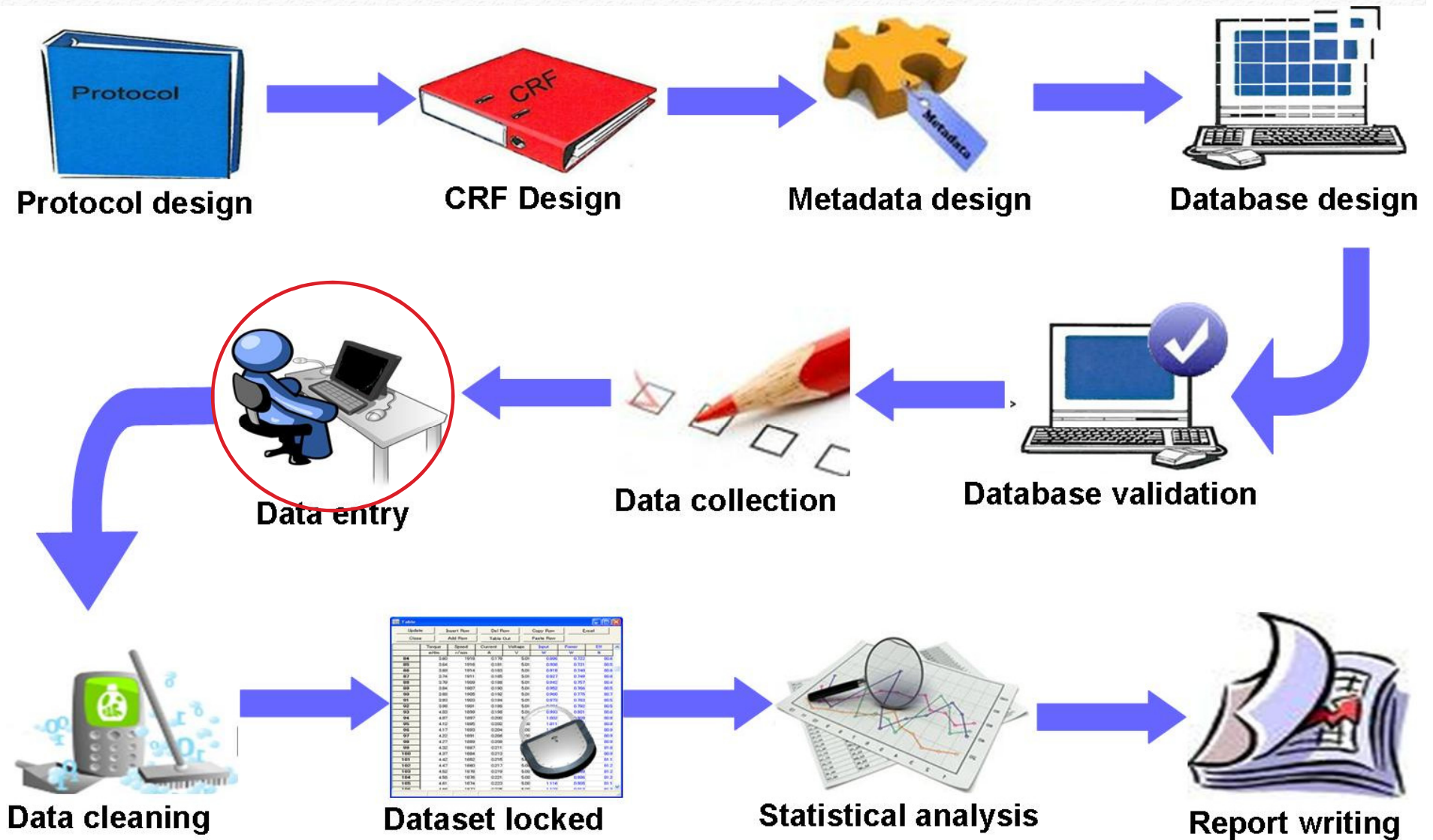
DATA LIFE CYCLE



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



Data entry following paper CRF collection

Browser: <https://iddmzwh3.its.unc.edu/CadetWeb/Medications.aspx> | Citrix Receiver | CadetWeb - Me... | landesbioscience... | Targeting chromat... | citeseenx.ist.psu.edu

CADETweb Name: ZZZ ZZZ MR#: 0000999999

Select Patient

- Consent
- Death
- Abstraction Dates
- HIV Care
- Risk Factors
- Allergies
- Vaccinations
- Screens
- Studies
- Hospitalizations
- Resistance Testing
- Illnesses
- Lab Results
- Medications
- Prophylactic Meds
- Other Meds
- Non-UNC Visits
- UNC Visits
- Vital Signs
- Survey Hx

SSO Navigation

- Log Out
- Cleaning Administration

ADD/UPDATE ENTRY:

Regimen:

Medication:

Start: (mm/dd/yyyy)

Start Trust:

End: (mm/dd/yyyy)

End Trust:

Dosage:

Route:

Study Med:

Side:

Reason:

Comments:

[Insert](#) [Cancel](#)

ART INFO UPDATING:

ART Start:

ART Start Trust:

HAART Start:

HAART Start Trust:

Overall Trust:

[Save Changes](#)

EXISTING MEDICATION ENTRIES:

Select	Reg.	Med Name	Start	ST	End:	ET	Study	Dose	Route	Side	Reason	Comments
Select 1	AZT	(Zidovudine, Retrovir)	06/08/1993	OK	06/15/1994	<6M		200mg tid	PO		Self Discontinuation (other)	Delete
Select 2	3TC	(Lamivudine, Epivir)	06/15/1994	<6M	11/14/1996	OK		150mg bid	PO		Other	Delete
Select 2	AZT	(Zidovudine, Retrovir)	06/15/1994	<6M	11/14/1996	OK						
Select 3	3TC	(Lamivudine, Epivir)	11/14/1996	OK	08/08/2000	OK						
Select 3	AZT	(Zidovudine, Retrovir)	11/14/1996	OK	08/08/2000	OK						
Select 3	Indinavir	(Crixivan)	11/14/1996	OK	08/08/2000	OK						
Select 4	Raltegravir	(Isentress, MK-0518)	04/12/2012	OK								
Select 4	Ritonavir Boost	(Norvir)	04/12/2012	OK								
Select 4	Truvada	(FTC, Tenofovir)	04/12/2012	OK								

ANTIRETROVIRAL MEDICATIONS:

(Includes: 3TC, FTC, AZT, d4T, ddC, ddI (Videx), ddI (Videx EC), Abacavir, Tenofovir, Combivir, Epizcom, Trizivir, Truvada, Delavirdine, Efavirenz, Nevirapine, Amprenavir, Atazanavir, Darunavir, Atripla, Fosamprenavir, Indinavir, Kaletra, Nelfinavir, Ritonavir, Ritonavir/boost (i.e., <1-400mg bid), Saquinavir (Fortovase), Saquinavir (Invirase), Tipranavir, T20, T1249, Maraviroc, Etravirine, Raltegravir, and Rilpivirine (Edurant). Complete - see medications list for more complete details.)

Overall ART Trust: _____ (i.e., OK, <1W, <2W, <1M, <2M, <3M, <6M, <1Y, <2Y, <3Y, <4y, <5y)

First ART Started Date: Year: _____ Month: _____ Day: _____ First ART Start Trust: _____ (OK, <1W, etc)

First HAART Started Date: Year: _____ Month: _____ Day: _____ First HAART Start Trust: _____ (OK, <1W, etc)

Reg.	Medication	Start Date	Start Trust	End Date	End Trust	Dosage	Route	Side Effect	Study	Reason for Discontinuation	Comments

*INCLUDE ALL YEARS.
*** IF Medication received as part of a research study then please list study in comments field.

Chart Abstraction Form Ver4 Dec2006_10meds2/7/0713


Page 6.

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

Study ID _____

Demographic Information Form (DI)
Revision 1



18. Does anyone in your household smoke? **DI1018**

1) Yes

2) No

19. Are you the biological parent of the child? **DI1019**

1) Yes

2) No

20. Are you: **DI1020**

1) Married

2) Living with a partner

3) Single (Go to Q. 22)

4) Divorced (Go to Q. 22)

5) Widowed (Go to Q. 22)

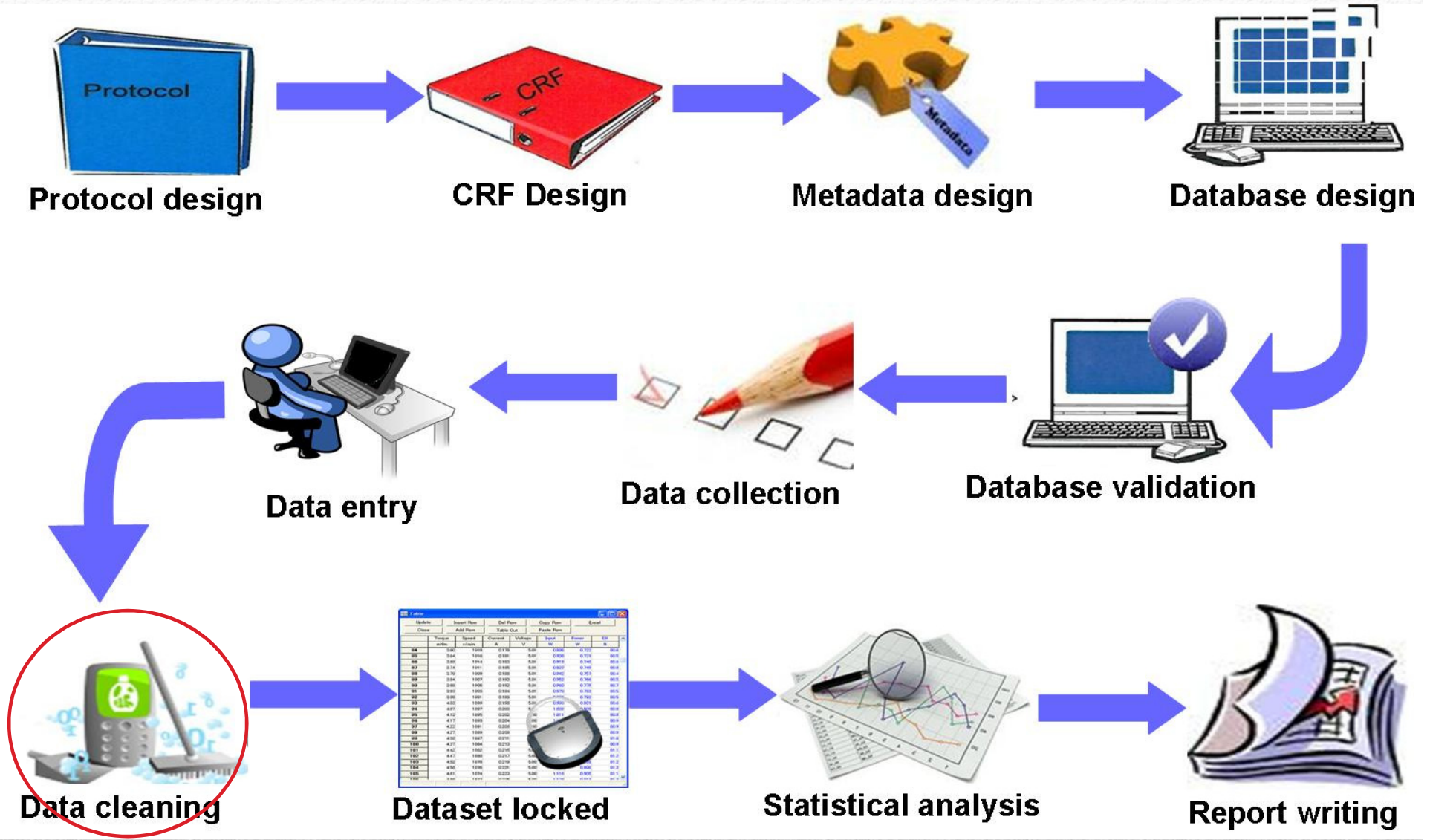
21. Is your spouse/partner currently employed? **DI1021**

1) Yes

2) No (Go to Q. 22)

21a. If yes, how many hours per week? _____ hrs. **DI1021a**

DATA LIFE CYCLE



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

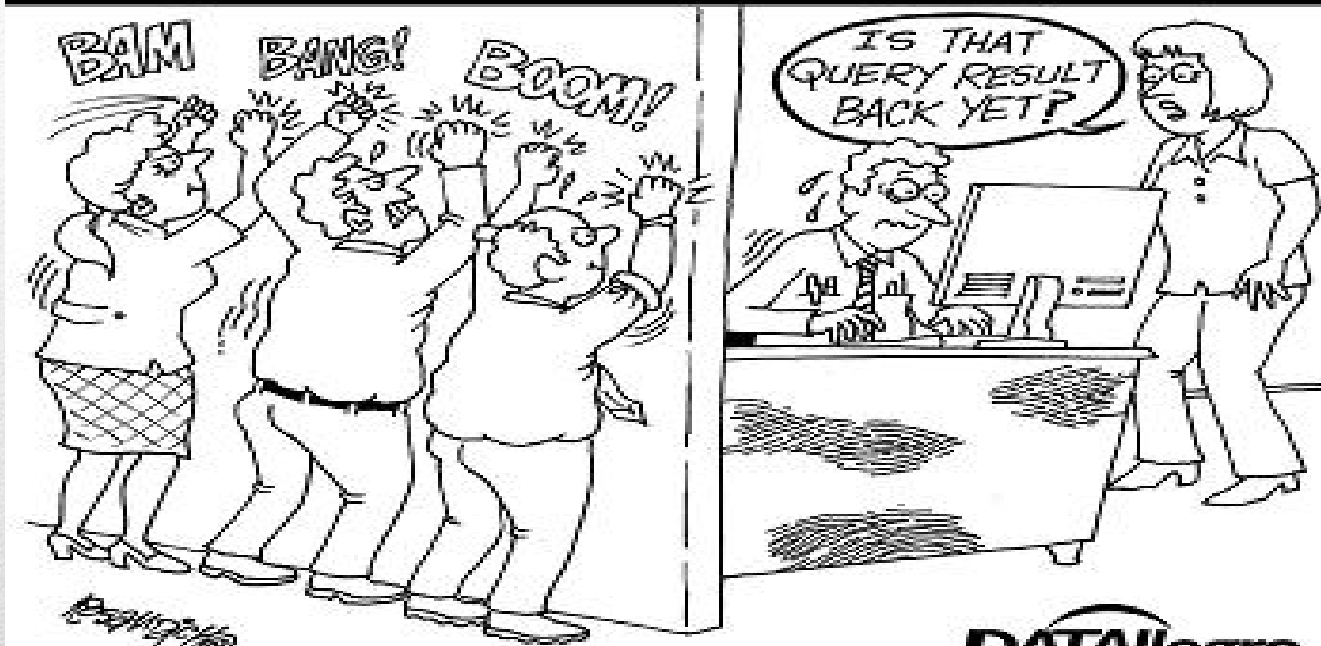
Continued...

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

F8 :

	A	B	D	E	F	G	J	K	L	M	N	O	P	Q	R
1	ID	test_date	CD_4	Medications	Med_StartDate	Med_endDate									
2	AAA	3/9/1997	131	AZT (Zidovudine, Retrovir)	4/6/1997	8/16/1998									
3	AAA	3/9/1997	131	Atazanavir (Reyataz)	1/20/2012										
4	AAA	3/9/1997	131	Atazanavir (Reyataz)	1/5/2012	1/20/2012									
5	AAA	3/9/1997	131	Truvada (FTC, Tenofovir)	4/21/2010										
6	AAA	3/9/1997	131	Ritonavir Boost (Norvir)	4/21/2010										
7	AAA	3/9/1997	131	Atazanavir (Reyataz)	4/21/2010	1/5/2012									
8	AAA	3/9/1997	131	Truvada (FTC, Tenofovir)	2/3/2009	3/11/2009									
9	AAA	3/9/1997	131	Ritonavir Boost (Norvir)	2/3/2009	3/11/2009									
10	AAA	3/9/1997	131	AZT (Zidovudine, Retrovir)	8/16/1998	9/10/1998									
11	AAA	4/6/1997	134	AZT (Zidovudine, Retrovir)	4/6/1997	8/16/1998									
12	AAA	4/6/1997	134	Atazanavir (Reyataz)	1/20/2012										
13	AAA	4/6/1997	134	Atazanavir (Reyataz)	1/5/2012	1/20/2012									
14	AAA	4/6/1997	134	Truvada (FTC, Tenofovir)	4/21/2010										
15	AAA	4/6/1997	134	Ritonavir Boost (Norvir)	4/21/2010										
16	AAA	4/6/1997	134	Atazanavir (Reyataz)	4/21/2010	1/5/2012									
17	AAA	4/6/1997	134	Truvada (FTC, Tenofovir)	2/3/2009	3/11/2009									
18	AAA	11/9/1997	132	AZT (Zidovudine, Retrovir)	4/6/1997	8/16/1998									
19	AAA	11/9/1997	132	Atazanavir (Reyataz)	1/20/2012	#VALUE!									
20	AAA	11/9/1997	132	Atazanavir (Reyataz)	1/5/2012	1/20/2012									
21	AAA	11/9/1997	132	Truvada (FTC, Tenofovir)	4/21/2010	#VALUE!									
22	AAA	11/9/1997	132	Ritonavir Boost (Norvir)	4/21/2010	#VALUE!									
23	AAA	11/9/1997	132	Atazanavir (Reyataz)	4/21/2010	1/5/2012									
24	AAA	11/9/1997	132	Truvada (FTC, Tenofovir)	2/3/2009	3/11/2009									
25	AAA	11/9/1997	132	Ritonavir Boost (Norvir)	2/3/2009	3/11/2009									
26	AAA	11/9/1997	132	Atazanavir (Reyataz)	2/3/2009	3/11/2009									
27	AAA	11/9/1997	132	Tenofovir (Viread)	4/24/2007	10/4/2008									
28	AAA	11/9/1997	132	Ritonavir Boost (Norvir)	4/24/2007	10/4/2008									
29	AAA	11/9/1997	132	Atazanavir (Reyataz)	4/24/2007	39725									
30	AAA	11/9/1997	132	3TC (Lamivudine, Epivir)	4/24/2007	39725									
31	AAA	11/9/1997	132	Tenofovir (Viread)	9/5/2006	39063									
32	AAA	11/9/1997	132	Saquinavir (Fortovase, Invirase)	38965	39063									
33	AAA	11/9/1997	132	Ritonavir Boost (Norvir)	38965	39063									
34	AAA	11/9/1997	132	Efavirenz (DMP266, Sustiva)	38965	39063									
35	AAA	11/9/1997	132	Abacavir (Ziagen)	38965	39063									
36	AAA	11/9/1997	132	Trizivir (AZT, 3TC, Abacavir)	38811	38965									

A Day in the Life of Databert



DATAlegro
(949) 330-7690
www.datalegro.com

Relational Databases

The screenshot shows the Microsoft Visual Studio Query Designer interface. The main window displays a query named 'Query1: Query(22043)'. The diagram shows two tables: 'APPDATA.MEDICATIONS' and 'WDATA.LABORATORY_FACT_TABLE'. They are connected by a line with a diamond at the end, indicating a full join relationship on the 'patient_key' column. The column grid below the diagram shows the following columns selected for output:

Column	Alias	Table	Output	Sort Type	Sort Order	Filter	Or...	Or...	Or...
regimen	Expr1	APPDATA....	<input checked="" type="checkbox"/>						
medication	Expr2	APPDATA....	<input checked="" type="checkbox"/>						
patient_key		APPDATA....	<input checked="" type="checkbox"/>						

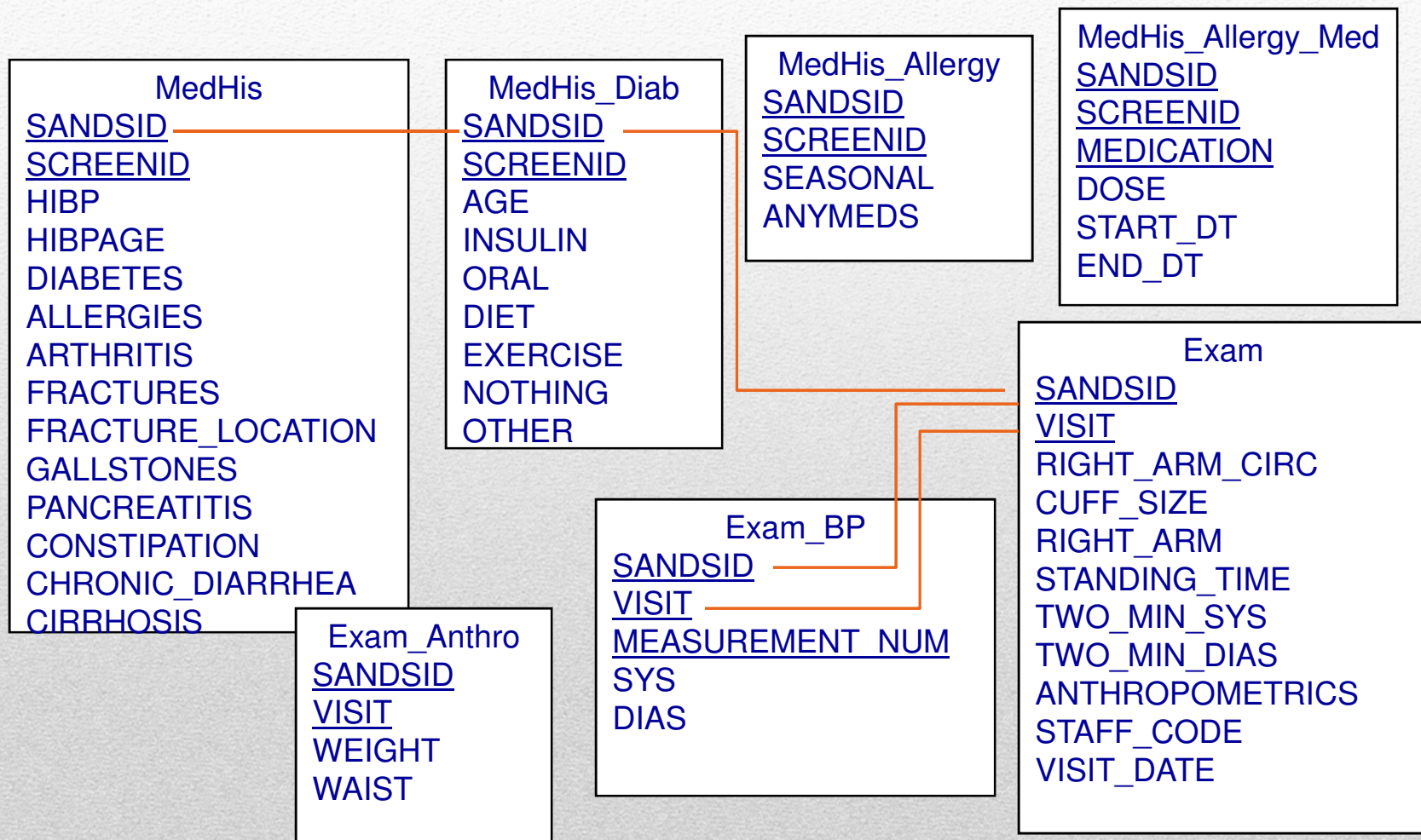
The SQL query editor shows the following query:

```
SELECT APPDATA.MEDICATIONS.regimen AS Expr1, APPDATA.MEDICATIONS.medication AS Expr2, APPDATA.MEDICATIONS.patient_key, APPDATA.MEDICATIONS.startDate, WDATA.LABORATORY_FACT_TABLE.test_number, WDATA.LABORATORY_FACT_TABLE.test_date FROM APPDATA.MEDICATIONS, WDATA.LABORATORY_FACT_TABLE WHERE APPDATA.MEDICATIONS.patient_key = WDATA.LABORATORY_FACT_TABLE.patient_key
```

The Output window at the bottom displays the following SQL query:

```
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



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
-



UNC
CFAR
Center for AIDS Research

UNC CFAR HIV/AIDS Clinical and Research Database:
<http://cfar.med.unc.edu/content/clinical-core>



Collaborative Studies Coordinating Center: <http://www2.csc.unc.edu/home/>

 **REDCap**TM
<http://tracs.unc.edu/>

Introducing...

EPI INFO 7

[Click here for more information](#)

<https://virtuallab.unc.edu/>

Data Management Courses:

- ODUM institute:

<http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=667>

Acknowledgment

- Sonia Napravnik
 - Joe Eron
 - Michael Hudgens
 - Katie Mollan
 - David Rosen
 - Carol Golin
 - Catherine Grodensky
 - Oksana Zakharova
 - Sam Stinnette
 - Christine Sun
-