Good Practices in Data Management for Statistical Analysis

Cathie Spino

For Environmental Statistics Discussion Series

With acknowledgement to Vicki L. Ellingrod (BIO588 lecture)

Evolution of CRFs

- Paper case report forms
  - Used to be ~75- 95% of all clinical trials
- Computerized systems (FDA 1997; *21 CFR 11 Electronic Records*)
  - Electronic image of paper CRF
  - Electronic data entry (web-based vs dynamic)
  - Electronic diaries
  - Mobile Phone data capture
  - Wearable devices
- Epidemiologic studies
  - Web-based questionnaires: the future in epidemiology? (2010 *Am J Epid)*
  - Mode of administration can influence results of psychosocial instruments

Example CRF – Study Medication

START WEB-ENTRY **ChiLDREN** Form S13B Steroid/Placebo

A1. Site/Study ID #: ___ ___ / ___ ___ ___ ___

A2. Discharge Date: ___ ___ / ___ ___ / ___ ___ ___ ___ (Month / Day / Year)

A3. Staff Initials: ___ ___ ___

To DCC ☐

STEROID TRIAL ONLY – ENTER ON WEB AT EACH CLINIC VISIT

SECTION B: STEROID/PLACEBO PRESCRIBED (Oral unless indicated otherwise)

Steroid/Placebo	Total Daily Dose	Start Date (mm/dd/yyyy)	End Date (mm/dd/yyyy)	Initials
B1. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B2. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B3. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B4. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B5. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B6. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B7. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B8. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B9. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __
B10. 1. ☐ Oral 2 ☐ Parenteral	_____ mg	__ __ / __ __ / __ __ __ __	__ __ / __ __ / __ __ __ __	__ __ __

EXAMPLE CRF – MATERNAL HISTORY

Database **BARC** **Form 05 Maternal Family Hx**

A1. Site/Study ID #: ___ ___ / ___ ___ ___ ___

C2. We want to know about any illnesses in members of your family that may be related to your child's illness. I will read you a list of illnesses. Please stop me and let me know if you or any members of your family, such as your other children, parents, grandparents or siblings, had a disease of this type.

Ask about all diseases that are listed	1. **N** = No 2. **Y** = Yes 3. **DK** = Don't know	**Which relative?** (Note all family members to whom this applies, using codes below*)
a. Liver disease while they were infants or children, such as:		
ai. Biliary atresia	☐ N ☐ Y ☐ DK	___ ___ , ___ ___ , ___ ___ , ___ ___, ___ ___
aii. Neonatal hepatitis	☐ N ☐ Y ☐ DK	___ ___ , ___ ___ , ___ ___ , ___ ___, ___ ___
aiii. Alpha-1-antitrypsin deficiency	☐ N ☐ Y ☐ DK	___ ___ , ___ ___ , ___ ___ , ___ ___, ___ ___
aiv. Alagilles syndrome	☐ N ☐ Y ☐ DK	___ ___ , ___ ___ , ___ ___ , ___ ___, ___ ___
av. Cystic fibrosis	☐ N ☐ Y ☐ DK	___ ___ , ___ ___ , ___ ___ , ___ ___, ___ ___
avi. Infant cholestasis	☐ N ☐ Y ☐ DK	

***Code for relatives of child: M** = Mother; **S1, S2, ...** = sisters starting from oldest; **B1, B2, ...** = brothers starting from oldest., **A1, A2, ...** = aunts (mother's sisters) starting from the oldest; **U1, U2, ...** = uncles (mother's brothers); **MGM** = maternal grandmother (mother's mother); **MGF** = maternal grandfather (mother's father).

EXAMPLE CRF - DEMOGRAPHICS

Definitions of Key Terms:

Race – an arbitrary classification based on physical characteristics; a group of persons related by common descent or heredity.

Ethnicity – an arbitrary classification based on cultural, religious, or linguistic traditions; ethnic traits, background, allegiance, or association.

Example CRFs: Demographics

- Write down your ideas to capture race
- Discuss with your partner
- We'll discuss together

CFLD PUSH STUDY

B3. What is the subject's racial background? **(check all that apply)**

- a. ☐ American Indian or Alaska Native
- b. ☐ Asian
- c. ☐ Black or African American
- d. ☐ Native Hawaiian or Other Pacific Islander
- e. ☐ White
- 55. ☐ Don't Know
- 9. ☐ Refused

Women's Health and Mood Screening Questionnaire

C2. What is your ethnic background? **(Check all that apply)**

a. ☐ Not Hispanic, Latino, or Spanish origin
b. ☐ Mexican, Mexican American, Chicana
c. ☐ Central American
d. ☐ South American
e. ☐ Puerto Rican
f. ☐ Cuban
g. ☐ Spanish
h. ☐ Other Hispanic, Latino, or Spanish origin

C3. What is your racial background? **(Check all that apply)**

a. ☐ White or European American
b. ☐ Black or African American
c. ☐ American Indian or Alaska Native
d. ☐ Asian Indian
e. ☐ Chinese
f. ☐ Filipino
g. ☐ Japanese
h. ☐ Korean
j. ☐ Vietnamese
k. ☐ Other Asian
l. ☐ Native Hawaiian
m. ☐ Guamanian or Chamorro
n. ☐ Samoan
o. ☐ Other Pacific Islander
p. ☐ Other race

PROBE STUDY

B4. What is *the infant's* racial background *(check all that apply)*?

a. ☐ American Indian or Alaska Native

b. ☐ Asian

c. ☐ Black or African American

d. ☐ Native Hawaiian or Other Pacific Islander

e. ☐ White

f. ☐ Other (Specify: _____)

g. ☐ DK

h. ☐ Refused

i. *If more than one response was chosen for B4a-B4g:* What would you say is *the infant's* primary racial background? *(choose only one)*

1. ☐ American Indian or Alaska Native

2. ☐ Asian

3. ☐ Black or African American

4. ☐ Native Hawaiian or Other Pacific Islander

5. ☐ White

6. ☐ Other (Specify: _____)

9. ☐ Refused

Environmental Pollution and Birth Outcomes in Mexico City

Place of Origin (delegation or municipality and corresponding state)

A. Where were you born?

_____

B. Where was your mother born?

_____

C. Where was your father born?

_____

Planned Enrollment Report

This report format should NOT be used for collecting data from study participants.

Study Title:

Domestic/Foreign:

	Ethnic Categories				
Racial Categories	Not Hispanic or Latino		Hispanic or Latino		Total
	Female	Male	Female	Male	
American Indian/ Alaska Native					0
Asian					0
Native Hawaiian or Other Pacific Islander					0
Black or African American					0
White					0
More Than One Race					0
Total	**0**	**0**	**0**	**0**	**0**

Comments:

Paper Case Report Forms

- Advantages
  - Time tested
  - Simple
  - Hard copies always available
- Disadvantages
  - Time delay from the events that are captured
  - Errors on remote entry (skip logic difficult)
  - Difficult and slow process for queries and error trapping

PEDSQL – PAPER FORM

*In the past **ONE month,** how much of a **problem** has your child had with …*

SiteID#:__ __ / __ __ __ __

PHYSICAL FUNCTIONING ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. Walking	0	1	2	3	4
2. Running	0	1	2	3	4
3. Participating in active play or exercise	0	1	2	3	4
4. Lifting something heavy	0	1	2	3	4
5. Bathing	0	1	2	3	4
6. Helping to pick up his or her toys	0	1	2	3	4
7. Having hurts or aches	0	1	2	3	4
8. Low energy level	0	1	2	3	4

PEDSQL – ECRF

Paediatric QOL Inventory PARENT RPT (TODDLERS) (QS3A)

*In the **PAST MONTH**, how much of a **problem** has your child had with...*

PHYSICAL FUNCTIONING *(problems with...)*

1.	1. Walking	List: PEDSQL1
2.	2. Running	List: PEDSQL1
3.	3. Participating in active play and exercise	List: PEDSQL1
4.	4. Lifting heavy things	List: PEDSQL1
5.	5. Bathing	List: PEDSQL1
6.	6. Helping to pick up his or her toys	List: PEDSQL1
7.	7. Having aches or pains	List: PEDSQL1
8.	8. Feeling tired	List: PEDSQL1

List:PEDSQL1

Label	Value
[Blank]	
Never	0
Almost Never	1
Sometimes	2
Often	3
Almost Always	4

Data Collection -- Exercise

- Look at the STUDY699 Form 03 Anthropometrics
- I provide information about BMI in the bottom half of the CRF for your information; it's not really part of the CRF
- What are the good features of this CRF?
- What are the bad features of this CRF?
- Discuss with your partner and write down 1-2 responses for each question above.

Tips for Data Collection

- During all phases of the study, ensure KEY data critical for interpretation of the trial are collected and are of high quality
  - Examples: patient information, history, inclusion, exclusion, dosing information, concomitant medicines, adverse events, primary safety and efficacy endpoints
- Collect raw data! (not calculated data)

SECTION B: ANTHROPOMETRIC MEASURES __ Not done

B1. Height ___ ___ ___ . ___ → 1. __ Inches 2. __ Centimeters

B2. Weight ___ ___ ___ . ___ → 1. __ Pounds 2. __ Kilograms

Tips for Designing Your CRF

- Design CRF to collect all data specified in protocol
  - Design CRF concurrently with protocol
  - Keep primary and secondary endpoints in mind
  - Avoid redundant data collection if possible– unless for data validity
  - Avoid data overkill
- THINK ABOUT HOW THE DATA WILL BE USED!

Data Collection -- Exercise

- PRINCESA Study Handout
  - Working backward from aim and analysis output, what data would you collect and how?
  - Think not only of the specific data fields, but about the forms on which the information would be collected. Would you have a certain form collected multiple times, at each visit? At baseline and at the end of the study? Only at the end of the study?

Tips for Designing Your CRF

- How you collect data influences the quality of the data:
  - Keep questions, prompts and instructions clear and concise
  - Minimize open ended/free text entries as much as possible
  - Maintain consistency of format/answers
  - Include- NA, Not applicable, No change, Other (see previous slide; e.g., Not done)
  - Prospectively think about how data will be coded and entered into a database for analysis!

Tips for Designing Your CRF

- Keep material organized in sections
- Design CRF to follow data flow from perspective of person completing it
- Pilot data collection instrument; revise if necessary
- Train all individuals as to how to use the CRF
- One resource: Society for Good Clinical Data Management
  - http://www.scdm.org/

Electronic Case Report Form (e-CRF)

- An auditable electronic record designed to record information required by the clinical trial protocol to be reported to the sponsor on each trial subject.

Reference: Guidance for Industry, Computerized Systems Used in Clinical Trials, April 1999

Designing eCRFs

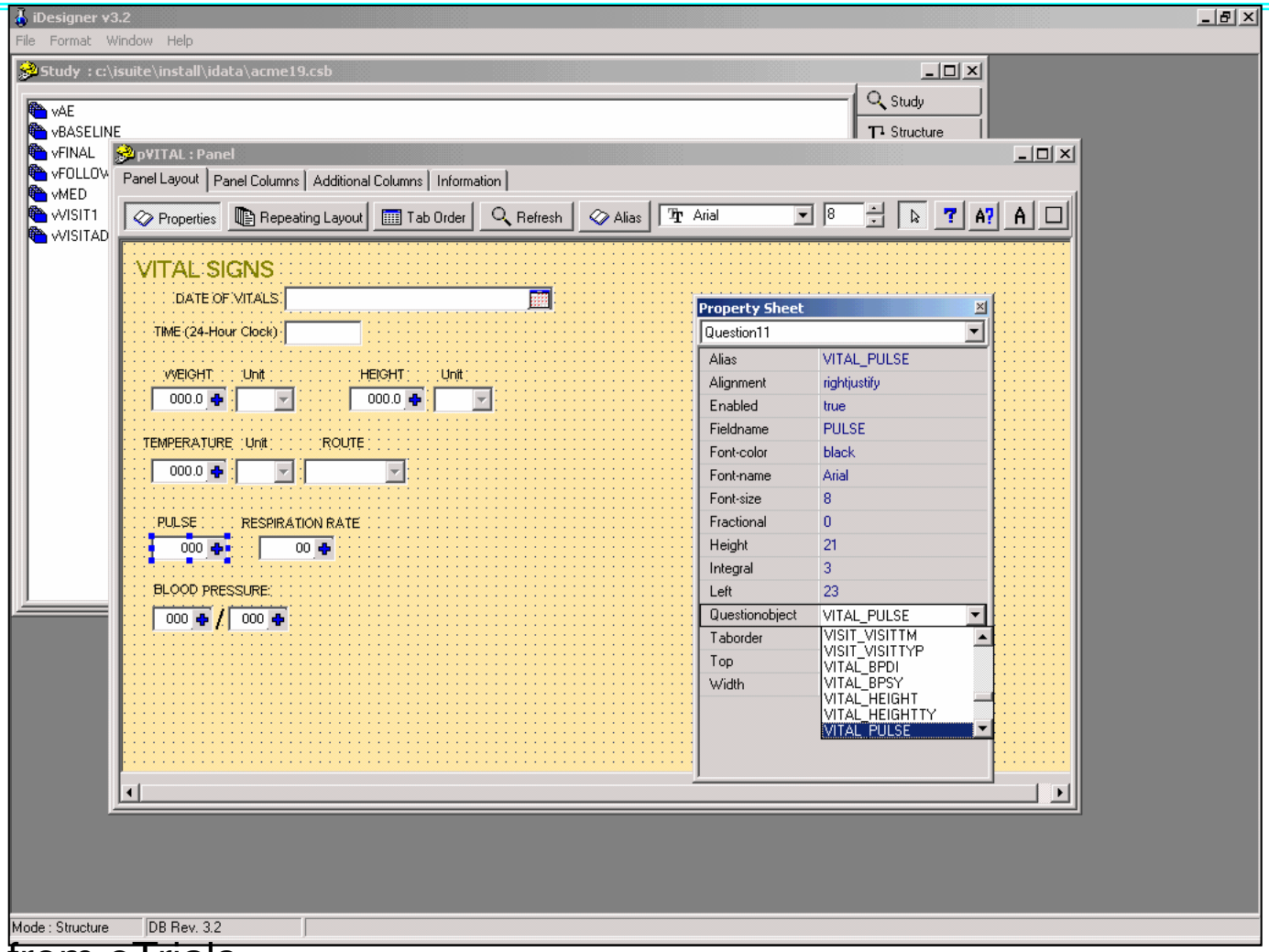


Image from eTrials

Easy Electronic Data Entry

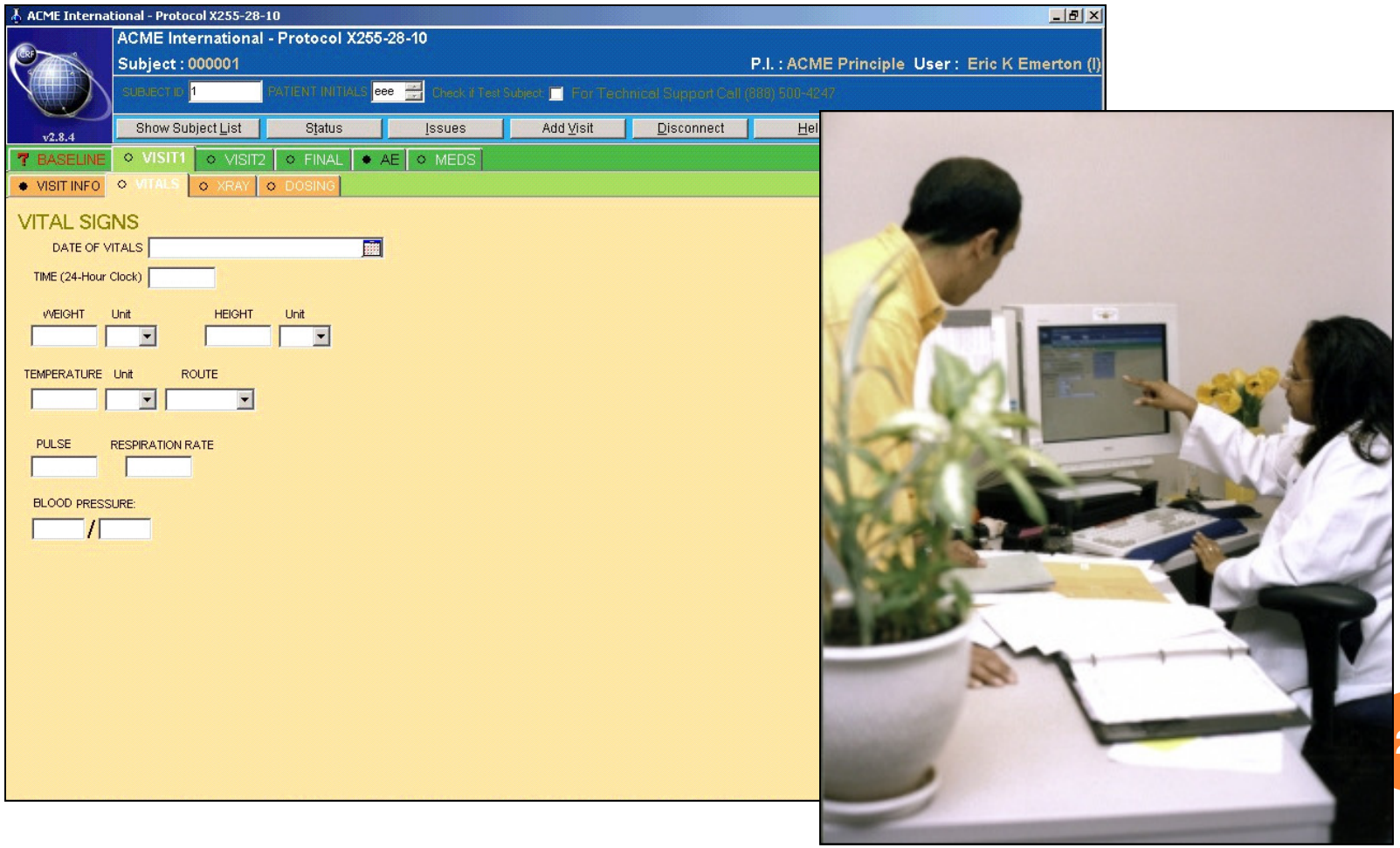


Image from eTrials

Electronic Data Capture in Academia

- If multicenter– may be worth investing in product
  - Prices vary depending on vendor, complexity of system, etc.
- Microsoft access (perhaps the academician's version of eCRF)

Data Management

- Why use a database?
  - Can create "pull down menus" which minimize data entry errors and facilitate consistent entry
  - Can have "front end" data checks as the coordinator is entering data
  - Relational database
  - Compatible with Excel, SAS, etc
  - Report generator

Computerized Systems System Features

- Prompts, flags, or other help features within the computerized system should be used to encourage consistent use of clinical terminology and to alert the user to data that are out of acceptable range.

Reference: Guidance for Industry, Computerized Systems Used in Clinical Trials, April 1999

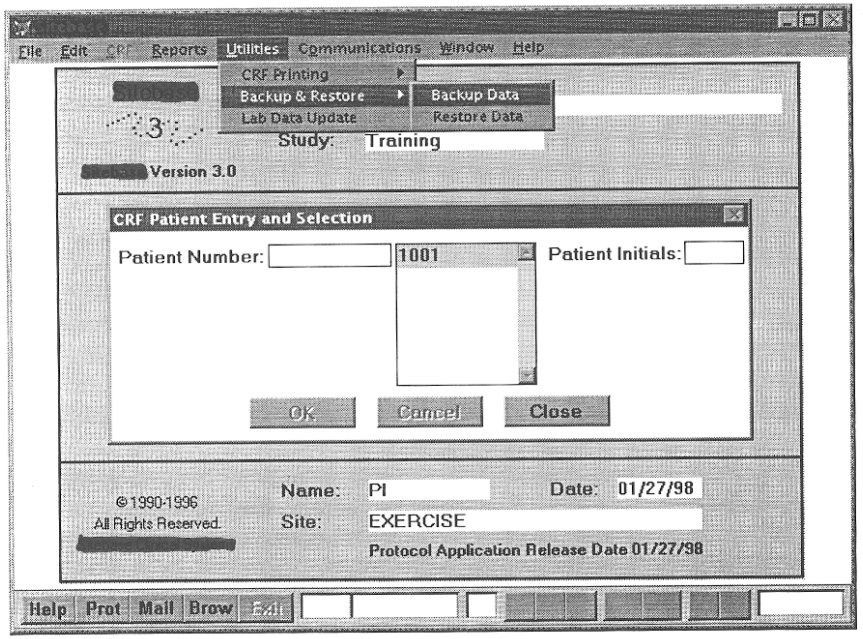
Computerized Systems Security / Electronic Signatures

- Password Protected:
  - *To ensure that individuals have the authority to proceed with data entry, the data entry system should be designed so that individuals need to enter electronic signatures, such as combined identification codes/passwords or biometric-based electronic signatures, at the start of a data entry session.*
  - *Individuals should only work under their own passwords or other access keys and should not share these with others.*
  - *Passwords or other access keys should be changed at established intervals.*
  - *When someone leaves a workstation, the person should log off the system. Failing this, an automatic log off may be appropriate for long idle periods. For short periods of inactivity, there should be some kind of automatic protection against unauthorized data entry. An example could be an automatic screen saver that prevents data entry until a password is entered.*

Reference: Guidance for Industry, Computerized Systems Used in Clinical Trials, April 1999

Computerized Systems Backup and Recovery

- Records should be backed up regularly in a way that would prevent a catastrophic loss and ensure the quality and integrity of the data.
- Backup records should be stored at a secure location. Storage is typically offsite or in a building separate from the original records.
- Backup and recovery logs should be maintained to facilitate an assessment of the nature and scope of data loss resulting from a system failure.



Reference: Guidance for Industry, Computerized Systems Used in Clinical Trials, April 1999

Three Major Problems in Data Collection

- Missing data
  - Not available
  - Poor collection (Not Done)
  - Careless record keeping
  - Poor CRF design
- Incorrect data
  - Unclear interpretation of item
  - Mislabeled specimens
  - Misrecorded
- Excess variability (Numerous sources of variability)
  - Factors that ↑ variability: vague definitions, inadequate methodology, lack of training, carelessness

Ensuring The Integrity of The Data

- Data Monitoring
- Data queries- internal/external
- Data entry- double enter; or enter and independently verify

Improving Data Integrity

- Pilot test data collection forms
- Train all personnel
- Standardization of study procedures (development of SOPs)
- Repeat measurements, blinded design, independent observation, supporting data for verification
- Computer checks to identify inconsistencies
- Data entry: double enter; or enter and independently verify
- Double data entry (to identify data entry errors)
- Monitoring/Surveillance System
- Data Queries: internal/external

REPRODUCIBILITY

- Clinical trials:
  - 2 adequate and well-controlled studies (21 CFR 314.126) for FDA
  - NIH sponsored studies in clinicaltrials.gov
    - Registered before enrollment
    - Results posted within 12 months of last subject last visit
- Epidemiologic studies
  - No laws that I'm aware of
  - Minimum standard: reproducibility
    - "independent investigators subject the original data to their own analyses and interpretations" (Peng, Dominici, Zeger. Reproducible epidemiologic research. *Am J Epid* 2006;163(9):783-789)
    - Data sets and software to be available

CONCLUSIONS

- Regardless of the type of study- identify, create, implement ways to ensure key data are collected and are accurate
- Develop and pilot test CRFs– collect what you need!!
- Electronic data collection, electronic patient recorded outcomes can facilitate data management
- Develop SOPs, train all personnel
- Maintain appropriate study records
- Accurately obtain and record data on CRFs
- Develop a quality control system (monitoring, auditing, error checks)