Validation of Derived Data via Cross Validation Methods

Emily Dai, Jim Chen, Cissy Tang, Janice Pogue, Dan Trottier

Population Health Research Institute, Mcmaster University and Hamilton Health Sciences Corp, Hamilton, Ontario, Canada

The Problems

- A data element may appear more than once in the eCRF
- To derive data based on calculation of previously collected data
- Continuation Studies needing subset of patient data from parent study
- Transfer of patients from a screening study to a registry type study

Within same study

- Examples of same data across different CRFs: Can they all be derived?
  - Patient Initials
  - Patient Age in the same year
  - Patient Weight, Height in the same month
  - Patient Initial Medication loading dose
  - Patient procedure method (PCI through radial or femoral) at baseline

Within same study

- Need to collect BMI
  - Can ask for it from clinic (they calculate and enter)
  - Can ask for component details and derive

Cross Studies

- Same study number but new schema/ setup
- Subset of existing patients going forward
- New CRFs going forward
  - ➢ New CRFs leverage some existing data
  - ➢ Want to pull data forward into new CRF plate from existing “older” study plate data

Cross Studies

- Transfer of Patient to New Study
  - ➢ Screening study with PHI data records
  - ➢ Patient is identified for specialized study
    - ❖ Want to transfer only certain patient record data to specialized study

So What Now?

- Can we just copy data from plate to plate?
- Or from study to study?
- Or derive data as needed?
- What do we need to worry about?

Points to Consider

- What does regulatory guidance say?
- What does the study protocol say?
- What are the risks to the patient?
- What are the risks to the study results?
- What are the mechanisms to do this?
- What verification or validation is needed?

Regulatory Guidance

- FDA 21 CFR Part 11

- Determine predicate rule requirements
- Assess the risk related to derived variable
- Implement appropriate Part 11 controls
  - Specify what's to be done (requirements)
  - Test/Validate correctness of what's being done
  - Control process (change management, documentation)

Regulatory Guidance

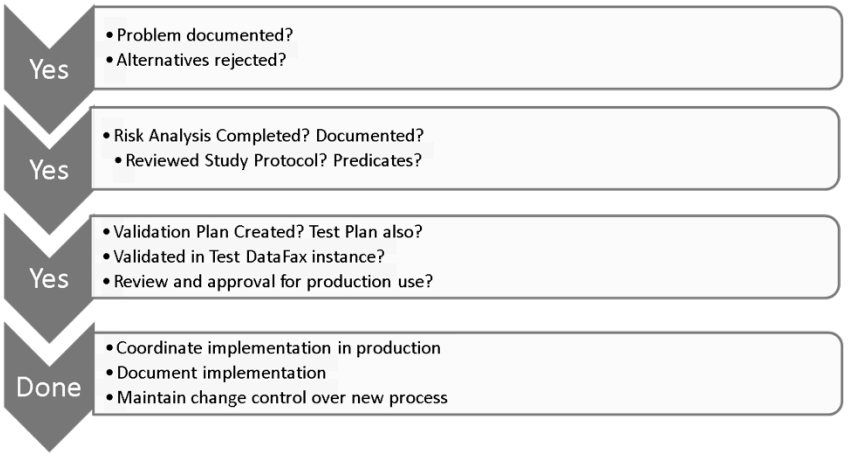
- FDA Guidance for Industry Computerized Systems Used in Clinical Investigations

- Study protocols
- SOP Source documentation and retention
- Internal security safeguards

Risks

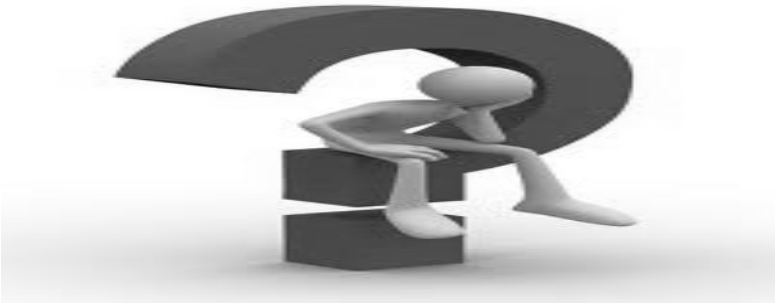
- NEVER hard code values in programming
  - ➢ If Patient initial on visit 1 equals “XYZ” then set Patient initial on visit 2 to “ABC”
- NEVER assume same data collection request is equivalent across CRFs(refer to study protocol)
  - ➢ Patient blood pressure on 3 CRFS for visit 2
- Data corruption during copy or calculation
- Audit trail clarity on “who” did “what”

Implementation Process



Options of data collection

- Manual Data Entry from original CRFs
- Edit Checks
- DFexport/DFimport



Edit Checks

- Pro
  - Programmatically efficient
  - Control Stays within DataFax System
  - Data stays within the DataFax System
  - Audit trail fully maintained
  - Lower cost
- Con
  - Introduces concern of 21 CFR Part 11
  - More coordinating time

DFexport/DFimport

- Pro
  - Simple to understand
  - Can be automated to process large data sets
  - Can be under the control of the study teams
  - Can be used to bring data into richer set of programming tools than edit check language
- Con
  - Introduces risk of data movement out of DataFax
  - Introduces risk of programming outside of DataFax
  - Audit trails do not show the originator. It shows only the person who does the import

DFexport/DFimport (cont)

- Con
  - Introduces concerns of 21 CFR Part 11
  - Requires conversion of export to import format
    - Will study teams use Excel to convert formats?
    - Will IT Department be leveraged to convert formats?

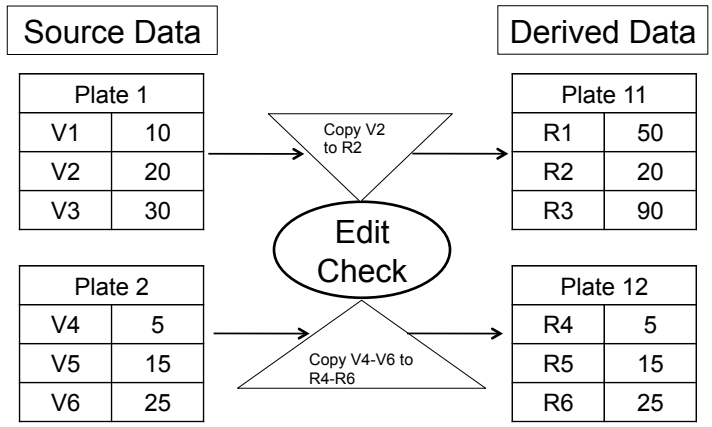
FDA requirements

- FDA issued 21 CFR part 11
  - Audit trails
  - System controls
- Guidance for industry computerized systems used in Clinical Investigations
- Standard operating procedures: create, modify and maintain data
- Training of Personnel

Cross Validation Procedures

- Standard edit check testing
  - Raw data testing (Cross plate checking)
  - Sample data testing
- R statistical programming language
  - R's language has a powerful, easy to learn syntax with many built-in statistical functions
  - R is open-source and runs on UNIX, Windows and Macintosh
  - Performs dataset comparisons easily

Example: Internal Cross Plate Data Transfer Using Edit Checks



*Source and destination fields must be defined the same

DFexport/DFimport Example

- Patient Screening Study 1 identifies candidates
- Candidate patient approved for Study 2
- Specified plate data is exported from study 1
- Exported data may need to be filtered (filtering criteria can be complicated)
- Exported plate data massaged into an import record
- Massaged record is tested/validated using Unix scripts (perl, awk, etc)
- Validated record is imported into Study 2
- Validation is done after import
- Scheduled jobs to do the export/import

How to maintain

- If the original study data management plan doesn't include this process initially, add an amendment
- Database is managed centrally, any modification and update of database setup will immediately inform the central personnel and proper procedure will be performed to ensure the integrity of database
- Risk assessments for all database changes

Are the risks acceptable?

- Should we even be doing this ?

Questions?