Challenge in Data entry and cleaning

CHAPTER 15. DOCUMENTS AND DATA MANAGEMENT

3 March 2011
Phnom Penh, Cambodia

Norio Yamada, RIT/JATA

Some part is contentious:
Data entry points are very large. Cleaning workload is also large.

Data collection and entry should be designed to reduce workload of them.
Data recording/entry format

• Documents need to be designed taking into account data entry

• Use multiple choice rather than typing numbers and characters in forms and in data entry as long as possible.
  – a. neg, b. scanty, c.1+, d. 2+, e. 3+, f. NA
  – not writing the results (e.g. 2+, ++, + +, )
Data entry

Household Register (Census Form) → Database-1
Survey Individual Form → Database-2 (Participants)
CXR Register → Database-3 (CXR)
Lab register → Database-4 (Lab)

Central Panel

S/C positive
CXR positive

Database-5 (Cases)
Data entry
May not recommended
More efforts and more mistake

Household Register (Census Form)

Survey Individual Form

CXR Register

Lab Register/Form

Database-1

Database-2 (Integrated)

Central Panel

S/C positive
CXR positive

Database-5 (Cases)

Filling the individual sheet

Individual Form including CXR and Lab results
How to ensure correct identification and linkage
Field Level Data Management

• **Field-level check is important.**
  – Usually it is difficult to go back to collect data after the team left a site.

• ID check in each step between forms/register and ID card
  – Interview
  – CXR (ID on film and ID on CXR register)
  – Sputum collection

• Check the form at check-out point of TB screening
  – Missing data, illegible
  – Inconsistency: e.g. Symptom/CXR/Eligibility for lab exam

• Review of forms after TB screening each day
Some redundancy of data recording/entry is necessary?(1)

- Records between different databases may be wrongly matched by only one matching variable (i.e. survey code) due to mistype/wrong recording.

- A few other variables are entered even if it is not used for comparison
  - E.g. Even if data of age/sex of participants is from individual survey sheet, age/sex is also recorded in lab forms/register and entered.
Some redundancy of data recording/entry is necessary? (2)

- From confidentiality viewpoints, recording personal identifiers (names) should be minimized.
- Name in questionnaire and registers (or forms) of laboratory and CXR [no need to entry in computer] might be helpful.
  - Persons in charge of treatment action check names of all the forms/registers to reduce unnecessary action to participants
  - Central panel needs to check matching for bacteriologically-confirmed cases to assess accuracy of recording/matching.
Some redundancy of data recording/entry is necessary?

- Both survey code and running number on both forms/records and films/Sputum cups might be helpful.
  - Wrong entry of survey code may happen.
  - Examination is carried out and recorded according to time order of attending TB screening. It is not easy to locate original forms/records in register only by survey code.
Using labels (Barcodes): probably reduce identification problems but enough preparation is required?
The End