TB prevalence survey issues related to data management

15.1 Essential documents:
- Structured and organized keeping of records for easy reference
- Storage by cluster
- Forms ordered by number
- Storage in secure, locked, adequate conditions

15.2.1 Data management – organizational asp.
- Do not underestimate workload of data management
- Data management start during protocol → system design
- Data entry should keep pace with survey →
  a) Collect missing data
  b) Perform checks → need someone to do this!
- Data management & analysis also tested in pilot → adjustment if needed
- Separate register for TB cases
- Fix case algorithm in protocol and stick to this
- Primary objective prevalence survey (for most countries) → SM+ prevalence

15.2.2 Data management – field procedures
- Separate census forms/register for adults and children (<15yrs) → blueprint for further survey (interviews, sputum) – example transcription
- Link between interview in HH and central level in cluster to decide who is suspect (duplicate interview/survey card)

15.2.x Data management – data manager
- Skills data manager: data manager should have skills to manage & validate data not just enter (merging sets, validations, checks & cross tabulations, etc).
- Assistance of data analyst
- Specific training if needed

15.2.x Data management – entry
- Software: try to use program country has experience with if possible but this should be a good data management program, EPI info is not recommended.
- Task force to provide list of suggested programs
- Use of PDAs by some countries, not an option for all. Need for built in checks, no double entry possible.
- Use form (like paper sheet) for easier data entry

15.2.x Data management – Quality control
- Double entry? Needed for all?
- i.e. lab data (error risk if >99% negative)
- Need for appropriate checks in data/between data files
- Primary versus secondary data
- Effect of error in case versus non case
Conclusions

- Minimize transcription of files/registers to avoid errors
- Record absent/refusals
- Data manager should coordinate (key) decision on data management with study coordinator
- At least all data management decisions should be well documented
- Design system in which people can not get “lost”, incorporate checks for this

Advice

- Data manager involved from the start
- SOPs need to include DM plan
- Data entry and data validation in pace with field activities
- Incorporate checks to validate data during the survey (and after) → list suggested checks?
- Have several registers (suspects and cases) for easy data retrieval
- Proper storage of data

Suggestion

- Task Force to design/create database infrastructure that countries can use as base
- Adapt for each country
- Provide internet based service with back up (for some countries)

Discussion by group /points raised afterwards:

a) It was stressed there is need to describe confidentiality issue clearly in protocol
b) What kind/percentage of error is allowed? This is not straightforward as it depends on which data. Between data error margin might differ. One error in a case has more influence than one error in a negative person. You would like to be 100% confident in your case classification but if the age of someone differs one year this might be less of a problem.
c) There is need for TA in data management. Countries do not realize the extend of it.
d) Long discussion on yes/no base data base infrastructure:
   Pros:  - Systematic base across countries
   - Option to build in checks
   - Better data quality?
   Cons:  - Hassle, do we really want to get into this?
   - Much effort/costs and not sure all countries use
   - Just advice proper TA
   Consensus: provide good TA to a country starting and use that as a base for all countries. Or use a good one as a base of the countries that have recently completed (Vietnam, Philippines, and Bangladesh). These countries might have a good example that could be used as a base for other countries.
e) Double entry vs. checks: no clear conclusion, depend on capacity country and data system and which type of data, no need for double entry for all data (see also above). Always need for good checks!