Efficient, Quality-assured
Data capture and analysis using EpiData

Zaw Myo Tun
Ajay M V Kumar

Special Acknowledgements
Dr Jens M Lauritsen (www.epidata.dk)
Hans L Rieder (www.tbrieder.org)
## Presentation components

<table>
<thead>
<tr>
<th>Part</th>
<th>Presenter</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ajay Kumar</td>
<td>Introduction to data capture and analysis</td>
</tr>
<tr>
<td>2</td>
<td>Zaw Myo Tun</td>
<td>Practical introduction to data form design and data capture with EpiData Manager and EpiData Entry Client</td>
</tr>
<tr>
<td>3</td>
<td>Ajay Kumar</td>
<td>Practical introduction to data analysis with EpiData Analysis</td>
</tr>
<tr>
<td>4</td>
<td>Ajay Kumar</td>
<td>Summary and conclusions</td>
</tr>
</tbody>
</table>
Introduction to data capture and analysis

Ajay Kumar

International Union Against Tuberculosis and Lung Disease
<table>
<thead>
<tr>
<th>Basic tenets for data entry and analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
</tr>
<tr>
<td>o</td>
</tr>
<tr>
<td>o</td>
</tr>
<tr>
<td>o</td>
</tr>
</tbody>
</table>
The right software for data capture
The right software for data capture

Not appropriate

- Spreadsheets: they are for calculations, not for data capture
- Analysis software (e.g. R, Stata, SAS, SPSS): for analysis, not suited for quality-assured data capture

Properties of the right software

- Non-proprietary: free and legal to distribute
- Small download size
- Small file size (e-mail transmission of files!)
- Stable, Simple, user-friendly, quick to learn
- Abundant data entry control options
- Simple validation procedure by record (not variables)
The right software for data analysis

Stratified analysis with Mantel-Haenszel procedure for adjustment

Kaplan-Meier survival estimation

KM adverse-free outcome probability by initial fluoroquinolone MIC

EpiData Analysis Graph
Thank you!

The next presenter is Zaw Myo Tun
Practical Introduction to
Data Form Design and Data Capture with
EpiData Manager and EpiData Entry Client

Zaw Myo Tun, Singapore
Outline

• Overview of EpiData Softwares
• EpiData Manager and EpiData Entry Client
• Double data entry and validation using the above softwares
Overview of EpiData softwares

• EpiData Entry (v 3.1)

• EpiData Analysis (v 2.2)

Freely downloadable from: www.epidata.dk
Overview of EpiData softwares

- EpiData Manager (v2.0.8.56)
- EpiData Entry Client (v2.0.7.22)
- EpiData Analysis (v2.2.2.183)

Freely downloadable from: www.epidata.dk
Why EpiData Manager?

- Users increasingly graphically oriented
- A need for multi-platform development (Windows, Mac)
- Use of Unicode UTF-8 $\rightarrow$ Non-latin characters
- Encryption and logging (GCP for medical data management)
- Control of user access to data
- Common “Engine” for EpiData software for data and metadata handling – same routine for different platform
- Should not be easy to change rule or structure during data entry $\rightarrow$ EpiData EntryClient
Ideal workflow for acquiring quality assured data

Create data entry form:
- Create fields
- Define fields and set entry rules

Double data Entry

Validating duplicate files

Feedback and correcting discrepancy

Quality assured dataset

Software: EpiData Manager
Personnel: e.g.
- Project manager
- Supervisor
- Principal Investigator

Software: EpiData Entry Client
Personnel: e.g. data clerk
EpiData Manager menu
EpiData Entry Client menu

Unique identifier

Participant's sex

How old are you?

Is this participant a TB case?
Data Documentation Sheet

- Field names
- Field label
- Field type
- Field values
- Value labels
Field Types

String field
- Any text (upper/lower cases, numbers)
- Upper case only string

Numeric field
- Integer
- Float (Decimal numbers)

Date field and time field

Boolean field (yes/no)
Four questions

- Unique identifier
- Participant’s sex
- How old are you?
- Is this participant a TB case?
### Example: Data Documentation Sheet

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Field Label</th>
<th>Field Type</th>
<th>Field Length</th>
<th>Field Value</th>
<th>Value Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Unique identifier</td>
<td>T</td>
<td>2</td>
<td>1,..., 80</td>
<td>NA</td>
</tr>
<tr>
<td>sex</td>
<td>Participant’s sex</td>
<td>I</td>
<td>1</td>
<td>1, 2, 9</td>
<td>Female, Male, Not recorded</td>
</tr>
<tr>
<td>age</td>
<td>Participant’s age</td>
<td>I</td>
<td>2</td>
<td>15,.....,90 (legal range) 99 (missing)</td>
<td>NA</td>
</tr>
<tr>
<td>case</td>
<td>Is this participant a TB case?</td>
<td>I</td>
<td>1</td>
<td>0, 1, 9</td>
<td>Non-case, Case, Not recorded</td>
</tr>
</tbody>
</table>
Practical Demonstration

• Designing a questionnaire in EpiData Manager
• Double data entry in EpiData EntryClient
• Validation in EpiData Manager
• Exploring other features of Epidata Manager and EpiData EntryClient (Subject to time availability)
Thank you!

The next presenter is Ajay Kumar
Practical introduction to data analysis with EpiData Analysis

Ajay M V Kumar

International Union Against Tuberculosis and Lung Disease
The Union South-East Asia Office
New Delhi, India
EpiData Analysis

- EpiData Analysis performs basic statistical analysis, graphs, and comprehensive data management
  - E.g: descriptive statistics, frequency, tables, box plot, ...
Program in EpiData Analysis

• Read data
• Browse
• Using the Editor
  – Using the Editor box
  – ALT+T short-cut key
  – F5
• Show the commands (F2) and the variables (F3)
Program in EpiData Analysis

• Describe, means
• Define numeric, text or date variable
• Frequency
• Tables
• Graph: bar, box plot, ciplot, ...

EpiData Analysis is available at:
http://www.epidata.dk
Summary and conclusions

Ajay Kumar

International Union Against Tuberculosis and Lung Disease
EpiData software covers >90 per cent of needs

- Free software with small download size:
  - EpiData Manager: 1.07 MB
  - EpiData Entry Client: 1.32 MB
  - EpiData Analysis: 1.87 MB

- Mastering EpiData Manager and Entry Client basics is a matter of a few hours

- Once the data documentation sheet is ready, building the entry form is swift

- Checks to control data entry can be simple, but with increasing experience sophistication is acquired to build in complexity to enhance user-friendliness

- Data entry is efficient with the proper design, providing powerful tools with metadata for numeric coding with explicit and unambiguous labels

- EpiData software is text-based, resulting in small file size: the sample set of 80 records sums up to 1.54 KB when zipped, convenient to e-mail

- Export of data files to Excel, Stata, SPSS and SAS

- EpiData Analysis gives power in data restructuring, tabulations, outbreak analysis, survival analysis, statistical process control charts, etc. It covers most needs, (purposefully) barring complex multivariate analyses.
You can do all your data entry and analyses with EpiData software, as required by a high impact factor journal, for instance this one:

Short, Highly Effective, and Inexpensive Standardized Treatment of Multidrug-resistant Tuberculosis

Armand Van Deun¹,², Aung Kya Jai Maug³, Md Abdul Hamid Salim³, Pankaj Kumar Das³, Mihir Ranjan Sarker³, Paul Daru³, and Hans L. Rieder¹,⁴

¹International Union Against Tuberculosis and Lung Disease, Paris, France; ²Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium; ³Damien Foundation Bangladesh, Dhaka, Bangladesh; and ⁴Institute of Social and Preventive Medicine, University of Zurich, Switzerland

Impact factor: 11.1

Am J Respir Crit Care Med 2010;182:684-92
Statement:

The Union collaborates closely with and supports the EpiData Association

Thank you for your interest