

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

Part	Presenter	Presentation
1	Ajay Kumar	Introduction to data capture and analysis
2	Zaw Myo Tun	Practical introduction to data form design and data capture with EpiData Manager and EpiData Entry Client
3	Ajay Kumar	Practical introduction to data analysis with EpiData Analysis
4	Ajay Kumar	Summary and conclusions

Introduction to data capture and analysis

Ajay Kumar

International Union Against Tuberculosis and Lung Disease

Basic tenets for data entry and analysis

- o The right software for data capture
- o Efficiency of data capture
- o Validation of captured data
- o Writing a program for analysis

The right software for data capture

The right software for data capture

Not appropriate

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

Properties of the right software

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

The right software for data analysis

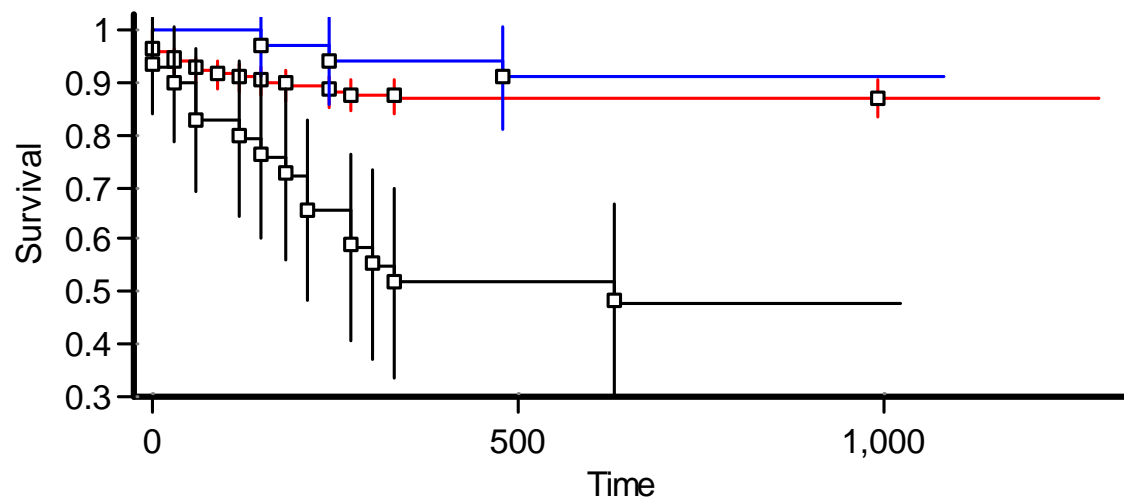
Binomial outcome by Patient's sex adjusted for Fluoroquinolone resistance level

	N = 501	N	OR	(95% CI)
Crude		501	1.78	(1.07-2.96)
Adjusted		501	1.61	(0.95-2.74)
Fluoroquinolone resistance level: Susceptible		439	1.52	(0.85-2.72)
Fluoroquinolone resistance level: Low-level resistance		33	2.50	(0.19-33.17)
Fluoroquinolone resistance level: High-level-resistance		29	2.06	(0.46-9.14)

Stratified analysis
with Mantel-Haenszel
procedure for
adjustment

KM adverse-free outcome probability by initial fluoroquinolone MIC

Kaplan-Meier survival
estimation



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 → 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 meta-data handling – same routine for different platform
- Should not be easy to change rule or structure during data entry → **EpiData EntryClient**

EpiData Entry



qes

rec

chk

EpiData Manager

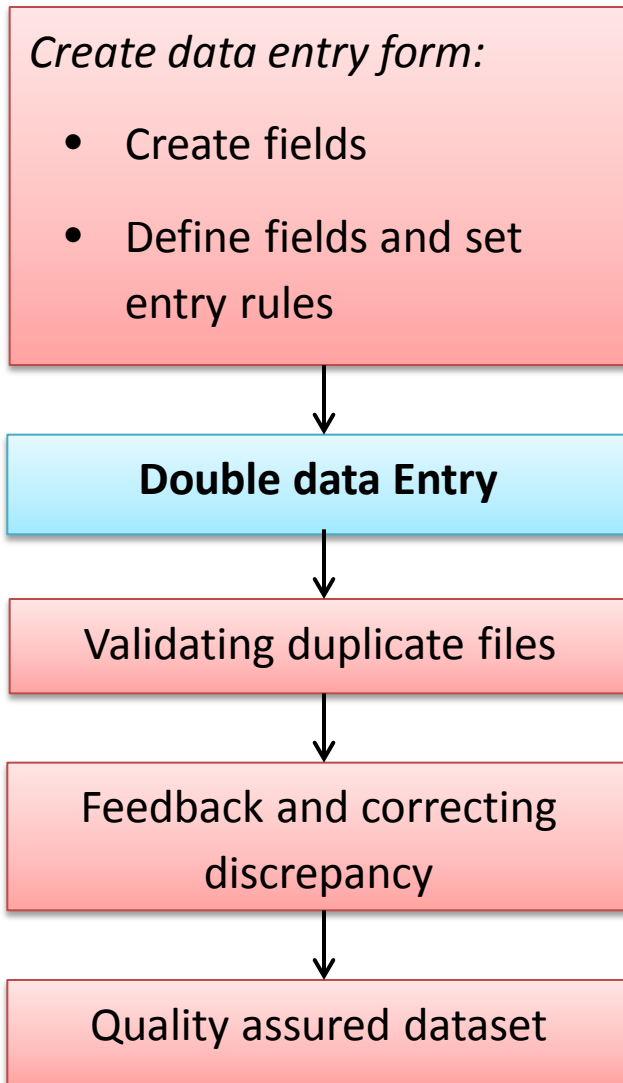


epx

EpiData Entry Client



Ideal workflow for acquiring quality assured data



Software: EpiData Manager


Personnel: e.g.

- Project manager
- Supervisor
- Principal Investigator


Software: EpiData Entry Client

Personnel: e.g. data clerk

EpiData Manager menu


 EpiData Manager (v2.0.8.56)


File Edit Document Tools Help


 Select Project

 Project Details

 Dataform

 Document

 Enter Data

 Export...

EpiData SAMPLE FILE V2.0 (february 2011)

This sample file was created with EpiData Manager. (see [Http://www.epidata.dk](http://www.epidata.dk))

Random data contained - only for testing and demonstration. See all sections.

Notice the different national characters used in the string field section

Integer Fields:

Integer Field (8 digits) Float Field (4.4 digits) **With range (50-250):**Integer field (3 digits)

Field with valuelabels:

Integer Field (2 digits) **Integet Set**Float Field (1.2 digits) **Float Set**String Field (3 characters) **String Set**

Date Fields:

Date Field (DD/MM/YYYY) Date Field (MM/DD/YYYY) Date Field (YYYY/MM/DD) Time Field (HH:MM:SS)

Auto Fields:

Auto Incrementing (4 digits) Date Field Auto (dmy) Date Field Auto (mdy) Date Field Auto (ymd) Time Field Auto

String fields:

String Field (20 characters)Language **String Field (40 characters), in language:**Sample text **String Field (40 characters), UPPERCASE in language:**Sample text

Jumps:

- jump field
- 0 = No Jump defined**
 - 1 = Skip Next (system missing)**
 - 2 = Exit Section (max missing)**
 - 3 = Jump to last field (2nd max)**
 - 4 = Save record (leave as is)**

With valuelabel **Integet Set**Float with second max **Float Set****Fields not in a section:**just a field Confirm-entry field The No-Enter field Last field...

EpiData Entry Client menu

 EpiData Entry Client (v2.0.7.22) - A.epx Version: 1 [Untitled Project]

File Edit Browse Data Goto Help

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)

Example: Simple Questionnaire

Four questions

- Unique identifier
- Participant's sex
- How old are you?
- Is this participant a TB case?

Example: Data Documentation Sheet

Field Name	Field Label	Field Type	Field Length	Field Value	Value Labels
id	Unique identifier	T	2	1,..., 80	NA
sex	Participant's sex	I	1	1 2 9	Female Male Not recorded
age	Participant's age	I	2	15,...,90 (legal range) 99 (missing)	NA
case	Is this participant a TB case?	I	1	0 1 9	Non-case Case Not recorded

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)



```
. read
Loading data C:\EPIDATA_COURSE\a.rec, please wait...
File name :C:\EPIDATA_COURSE\a.rec
Combined set labs A, B, C, and D
Fields: 8 Total records: 75 Included: 75
```

```
. freq sex
```

Examinee's

sex		N
Female	33	
Male	42	
Total	75	

```
. freq sex /v1
```

Examinee's

sex		N
1 Female	33	
2 Male	42	
Total	75	

Result Window
(Viewer)

```
G:\Conference 2010\Epidata_analysis_example\introduction.pgm - ...
File Edit Insert History Run Window Help
1 cls
2 close
3 logclose
4
5 read "a.rec"
6 freq sex
7 freq sex /v1
```

Editor (F5)

Line:7 Col:13 Insert

- Read & start
- View Data
- Analysis
- Graphs
- Save Output, Clear Screen
- Generate/Change Variables
- SPC graphs
- Save-Sort-Edit Data
- Clean Up and Stop
- Information
- Dis

Commands (F2)

```
labcode S Laboratory code
regdate D Registration date
age I Examinee's age in years
sex I Examinee's sex
reason I Examination reason
res1 F Result of specimen 1
res2 F Result of specimen 2
res3 F Result of specimen 3
```

Variables (F3)

```
run C:\Program Files\EpiData\epidatostat.ini
set echo=on
set viewer font size =10
set window font size =10
set editor font size =10
set viewer font name ="Verdana,Courier"
set display variables=OFF
set display
set output
cd C:\EPIDATA_COURSE
set output folder="C:\temp"
set language=english
set echo=ON
view "C:\Program Files\EpiData\docs\EN\start.htm"
```

History (F7)

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

- o Free software with small download size:
 - EpiData Manager: 1.07 MB
 - EpiData Entry Client: 1.32 MB
 - EpiData Analysis: 1.87 MB
- o Mastering EpiData Manager and Entry Client basics is a matter of a few hours
- o Once the data documentation sheet is ready, building the entry form is swift
- o Checks to control data entry can be simple, but with increasing experience sophistication is acquired to build in complexity to enhance user-friendliness
- o Data entry is efficient with the proper design, providing powerful tools with metadata for numeric coding with explicit and unambiguous labels
- o 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
- O Export of data files to Excel, Stata, SPSS and SAS
- O 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 Deun^{1,2}, Aung Kya Jai Maug³, Md Abdul Hamid Salim³, Pankaj Kumar Das³, Mihir Ranjan Sarker³, Paul Daru³, and Hans L. Rieder^{1,4}

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