Module 4: Data analysis and presentation
Six steps in the IR process

1. Contextualize challenges
2. Develop a proposal
3. Plan and conduct project
4. Analyse and present data
5. Disseminate research findings
6. Monitor and evaluate the project
What are the differences between quantitative and qualitative research?

- Research questions
- Methodological differences
- Data analysis
Comparing qualitative and quantitative approaches

<table>
<thead>
<tr>
<th></th>
<th>Qualitative</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social theory</strong></td>
<td>Action</td>
<td>Structure</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>Observation, interview</td>
<td>Experiment, survey</td>
</tr>
<tr>
<td><strong>Question</strong></td>
<td>What is x? How? Why?</td>
<td>How many xs?</td>
</tr>
<tr>
<td></td>
<td>(classification)</td>
<td>(enumeration)</td>
</tr>
<tr>
<td><strong>Reasoning</strong></td>
<td>Inductive</td>
<td>Deductive</td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
<td>Theoretical</td>
<td>Statistical</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>Validity</td>
<td>Reliability</td>
</tr>
</tbody>
</table>

Pope and Mays (1995). Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ*: 311; No. 6996
Module 4a: Quantitative data analysis and presentation
Presentation outline

Expected outcomes
Key concepts
  Data analysis plan
  Quantitative data analysis
  Data management
Learning Objectives & Expected outcomes

Able to:
- Describe data analysis planning processes
- Understand appropriate statistical measures
- Understand data management approaches
- Appreciate the importance of tailored / audience sensitive data presentation
Key concept 1: Data analysis plan

Designing analysis for use in IR aims to:

- Understand the implementation processes
- Communicate the implementation process to stakeholders

"Emphasis on simplicity and interpretability"
Key concept 1: Data analysis plan

Designing analysis for use
Different stakeholders need different information:
  Lay people?
  Community leaders?
  Local government/health service leaders?
  Civil society and media personnel?
  National policy-makers?

“Emphasis on simplicity and interpretability”
Designing analysis by purpose focuses on the objective of the analysis:

- Effectiveness
- Efficiency
- Equity
- Sustainability
Key concept 1: Data analysis plan

Data presentation formats

Data reporting should be presented in both textual and visual formats, such as:

- Tables
- Diagrams
- Graphs
- Infographics
- Maps
## Provider education expressed as frequency table

<table>
<thead>
<tr>
<th>Level of education of private providers</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illiterate</td>
<td>106</td>
</tr>
<tr>
<td>Basic literacy</td>
<td>74</td>
</tr>
<tr>
<td>Primary school certificate</td>
<td>57</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>11</td>
</tr>
<tr>
<td>Higher level qualification</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>250</strong></td>
</tr>
</tbody>
</table>
Joint frequency distributions for two or more variables

<table>
<thead>
<tr>
<th>Highest level</th>
<th>Men</th>
<th>Women</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illiterate</td>
<td>42</td>
<td>64</td>
<td>106</td>
</tr>
<tr>
<td>Basic literacy</td>
<td>45</td>
<td>29</td>
<td>74</td>
</tr>
<tr>
<td>Primary school certificate</td>
<td>32</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>8</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Higher level qualification</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>128</strong></td>
<td><strong>122</strong></td>
<td><strong>250</strong></td>
</tr>
</tbody>
</table>
## Provider education presented as proportion, percentage and cumulative %

<table>
<thead>
<tr>
<th>Level of education</th>
<th>Proportion</th>
<th>Percentage</th>
<th>Cumulative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illiterate</td>
<td>0.424</td>
<td>42.4</td>
<td>42.4</td>
</tr>
<tr>
<td>Basic literacy</td>
<td>0.296</td>
<td>29.6</td>
<td>72.0</td>
</tr>
<tr>
<td>Primary school certificate</td>
<td>0.228</td>
<td>22.8</td>
<td>94.8</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>0.044</td>
<td>4.4</td>
<td>99.2</td>
</tr>
<tr>
<td>Higher level qualification</td>
<td>0.008</td>
<td>0.8</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.000</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
<tr>
<td>Highest level</td>
<td>Men</td>
<td>Women</td>
<td>All</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>Illiterate</td>
<td>39.6</td>
<td>60.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Basic literacy</td>
<td>60.8</td>
<td>39.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Primary school certificate</td>
<td>56.1</td>
<td>43.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>72.7</td>
<td>27.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Higher level qualification</td>
<td>50.0</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>51.2</strong></td>
<td><strong>48.8</strong></td>
<td><strong>100.0</strong></td>
</tr>
<tr>
<td>Highest level</td>
<td>Men</td>
<td>Women</td>
<td>All</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Illiterate</td>
<td>32.8</td>
<td>52.5</td>
<td>42.4</td>
</tr>
<tr>
<td>Basic literacy</td>
<td>35.2</td>
<td>23.8</td>
<td>29.6</td>
</tr>
<tr>
<td>Primary school certificate</td>
<td>25.0</td>
<td>20.5</td>
<td>22.8</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>6.3</td>
<td>2.5</td>
<td>4.4</td>
</tr>
<tr>
<td>Higher level qualification</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
Provider education expressed as a bar chart

Level of education

- Illiterate
- Basic literacy
- Primary school certificate
- Secondary school certificate
- Higher level qualification
Provider education presented as a pie chart

- 4% Secondary school certificate
- 23% Primary school certificate
- 30% Basic Literacy
- 1% Higher level qualification
- 42% Illiterate
Line graph for trend analysis

Average *Ae. Aegypti* population per week in 5 field study sites
In what other format(s) can this data set be presented?
Map for spatial distribution

Community survey and water testing sample distribution
20th century death

Source: Information is beautiful
url: http://www.informationisbeautiful.net/visualizations/20th-century-death/
Potential tax revenue from drugs

Drug Deal?
Potential tax revenue from legalising drugs in the UK

Most users
- Cannabis "Skunk" £461m tax revenue
- Cannabis Resin £381m
- Cocaine £162m
- Ecstasy £89m
- Amphetamine £112m
- LSD £40m
- Heroin £212m
- Crack £474m

Least users

£5bn
Total revenue from legalization of drugs

+ £1.8bn tax
+ £0.3bn VAT
+ £1.7bn savings from societal costs
+ £1.0bn income tax from working users
+ £0.5bn inflation adjustment
- £0.3bn additional societal costs

£9bn
Government income from alcohol

£5bn
Shortfall in university funding 2012

£10bn
Government income from tobacco

£7bn
Bailout for Irish republic

David McCandless & Joe Swainson
v 1.0 // Dec 10 // informationisbeautiful.net

source: Independent Drugs Monitoring Unit (idmu.co.uk)
data: bit.ly/drug_deal

Source: Information is beautiful site
http://www.informationisbeautiful.net/visualizations/drug-deal-potential-tax-revenue-from-legalized-narcotics/
20 oz cola = 16 tsp sugar
Interactive graph

Source: Gapminder

http://www.gapminder.org/world-offline/
Whiteboard animation

Source: Eliminate Dengue Program
URL: http://www.eliminatedengue.com
In your project, discuss the results of the study that you need to disseminate and format of data presentation you will use for different stakeholders.
Key concept 2: Quantitative data analysis

Depending on research question:
Descriptive vs analytic study?
Analytic study, what to find?
Association
Causality
Statistical difference
Variables in quantitative analysis are usually classified by their level of measurement:

- Rational
- Interval
- Ordinal
- Nominal
Key concept 2: Quantitative data analysis

Descriptive statistics
- Distributions and summary measures
- Defining intervals for frequency distributions
- Frequency distribution and summary statistics
- Measures of variation
Distributions and summary measures

Advantages of frequency distributions:

useful for all types of variables

easy to explain and interpret

presented graphically and in different formats
Key concept 2:
Quantitative data analysis

Defining intervals for frequency distributions

Constructing a frequency distribution requires a choice of intervals:
- Ordinal
- Interval
- Rational

Two conflicting objectives when determining intervals:
- Limiting the loss of information
- Providing a simple, interpretable and useful summary
A powerful and robust form of analysis. Summary statistics usually focus on: overall location of a distribution or extent of variation within a population.
Key concept 2: Quantitative data analysis

Use of mean or median

Mean – the average value

Median – the value in the middle
Use of mean or median

Normal distribution

Skewed distribution
Key concept 2: Quantitative data analysis

Measures of variation

How much variability?
Low variability
High variability
Key concept 2:
Quantitative data analysis

Measures of variation

Choices of measures
- Variances
- Standard deviations

Alternative measures
- Quartiles: divide data into four quarters (Q1 to Q4) – 25% in each
- Percentiles: divide the data into two parts
Key concept 2: Quantitative data analysis

Analytical statistics
- Group comparison
- Association
- Causality
### Key concept 2: Quantitative data analysis

<table>
<thead>
<tr>
<th>Measurement scale</th>
<th>Assumption of distribution</th>
<th>Type of group</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal or Ordinal</td>
<td>-</td>
<td>Independent</td>
<td>Chi square test</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Paired</td>
<td>Sign test</td>
</tr>
<tr>
<td>Interval or Ratio</td>
<td>Normally distributed</td>
<td>Independent</td>
<td>Independent test</td>
</tr>
<tr>
<td></td>
<td>Not normally distributed</td>
<td>Paired</td>
<td>Paired test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Independent</td>
<td>Median test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paired</td>
<td>Wilcoxon</td>
</tr>
</tbody>
</table>
Finding association

Pearson correlation
- Ratio/interval scale
- Normal distribution of data

Rank correlation
- Ratio/interval scale
- Non normal distribution of data

Chi Square
- Categorical data
Key concept 2: Quantitative data analysis

Causality (regression)

Linear regression
- Continuous variable of both independent and dependent variable
- Normal distribution of data

Logistic regression
- Dichotomous dependent variable
- Continuous and categorical independent variable
Key concept 2: Quantitative data analysis

Causality (regression)

Cox proportional hazard model

Time-dependent outcome (survival model)

Continuous and categorical independent variable
Risk and odds used interchangeably, but not the same

 Reduction in risk is not equivalent to reduction in odds

Measures of risk

‘Risk’ and ‘odds’ used interchangeably, but not the same

‘Reduction in risk’ is not equivalent to ‘reduction in odds’
Measures of risk:
The ‘denominator problem’

Risk calculation requires calculation of the population at risk

Provide the estimates of both the numerator and denominator alongside any proportion, percentage or risk estimate
Sub-group analysis

The outcomes of an intervention may differ among sub-groups.

‘Data mining’ is useful to formulate new hypotheses but requires great caution in IR.
In your project, discuss the data analysis that you will do and identify whether the data you are collecting is suitable for the type of analysis you plan.
Key concept 3: Data management

Principle of data management
Data management and study phase
Key concept 3: Data management

Data quality and integrity

Data should be:
- High quality
- Reliable

“No study is better than the quality of its data”
Key concept 3: Data management

Prior to data collection process

ID number
Flow of data collection and handling process
Protocol for quality control
  Checking interviewee response
  Re-interview process
Electronic database development
SOP for data entry process
Key concept 3: Data management

Data collection process

- Data collection supervision
- Questionnaires/data collection forms storage management
- Checking data entry process
Key concept 3: Data management

Post-data collection process

- Checking database consistency
- Data cleaning
- Data coding
In your project proposal, discuss how to improve the quality of their data management system.
Conclusion

Start from the end

Plan your data analysis according to stakeholders’ need for information

Use appropriate statistical tools according to the information needed

Manage your data to ensure the validity of the data collected