



Research Methodology

Assoc. Prof. Dr. Jamalludin Ab Rahman
B.Med.Sc., MD, MPH
Department of Community Medicine
Kulliyah of Medicine
International Islamic University Malaysia

The lectures will answer the following questions

- ▶ Why we do research?
- ▶ How to plan a scientific research?
- ▶ What is the best design for my research?
- ▶ How to select subjects for research?
- ▶ How to plan for data collection?
- ▶ How to manage & analyse data?
- ▶ How to report the findings?



الجامعة الإسلامية
International Islamic University
Malaysia

Research Methodology Workshop
22-24 April 2014



Research Methodology Workshop

for Oral Health Division, JKNP
22-24th April 2014

Programme

- ▶ Day 1 – Reason, rationale & conceptual framework; Study design & Sampling plan
- ▶ Day 2 – Data collection & data dictionary; Statistical analysis & dummy table
- ▶ Day 3 – Writing proposal & report

Research Methodology Workshop
22-24 April 2014



المنظمة
الصحية
السعودية

Part 1 – Reasons & Rationale



UNIVERSITY OF
BIRMINGHAM

Why we do research?

- ▶ To answer curiosity
- ▶ To solve problem
- ▶ To find alternative
- ▶ To fulfil academic requirements
- ▶ To get promotion



UNIVERSITY OF
BIRMINGHAM

When the topic is important?

- ▶ When it fulfils the need
- ▶ Humankind need
- ▶ Country need
- ▶ Service need
- ▶ But never a personal need



المملكة العربية السعودية
Ministry of Health

Research Methodology Workshop
22-24 April 2014

Type of research

- ▶ Descriptive – e.g. prevalence study
- ▶ Discovery – e.g. new biomarker, new pathway (mostly lab study)
- ▶ Proof causality (hypothesis testing) – e.g. cause/causes of a disease or abnormality, reason for certain behaviour, clinical trial (a drug is better than others) etc



المملكة العربية السعودية
Ministry of Health

Research Methodology Workshop
22-24 April 2014

The steps

1. Organise research idea
2. Understand the problem → do literature review, construct conceptual framework
3. Formulate objective
4. Know your target population
5. Choose best study design
6. Choose best sampling method
7. Calculate sample size
8. Determine variables to collect → prepare data dictionary
9. Validate research instrument
10. Plan data collection, plan for quality control
11. Plan statistical analysis, prepare dummy table
12. Estimate & secure budget
13. Get approval
14. Collect data, monitor quality
15. Analyse data
16. Report finding



الجامعة الأردنية
الأردن
الجامعة الأردنية

Research Methodology Workshop
22-24 April 2014

Literature review

- ▶ Search strategically
 - ▶ Identify important keywords
 - ▶ Identify authorities in the subject matter
 - ▶ Observe the dates
- ▶ Manage bibliographic well – use software (e.g. EndNote, Mendeley, Papers)



الجامعة الأردنية
الأردن
الجامعة الأردنية

Research Methodology Workshop
22-24 April 2014

Part 2 – Conceptual Framework



المنظمة الوطنية
للحماية من الأمراض

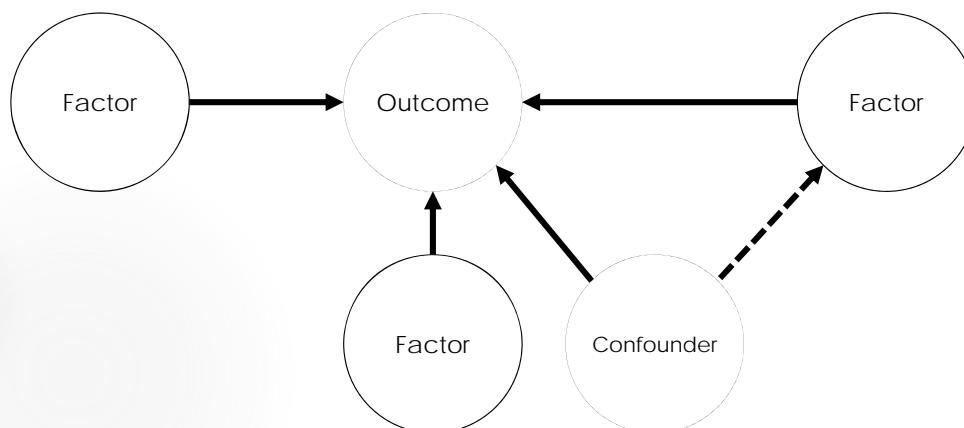
Build the concept

1. Main outcomes
2. Explanatory (exposures, factors) variables
3. Confounding variables



المنظمة الوطنية
للحماية من الأمراض

Conceptual idea

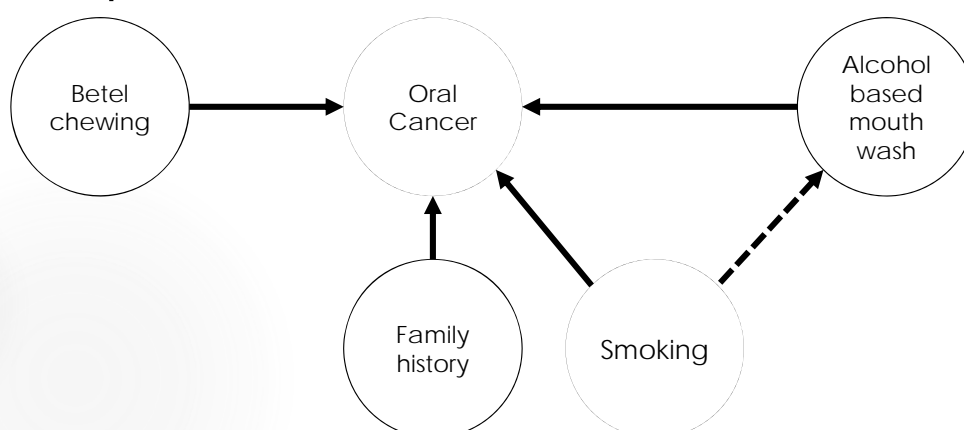


Research Methodology Workshop
22-24 April 2014



الجامعة الأردنية
Amman, Jordan

Example 1

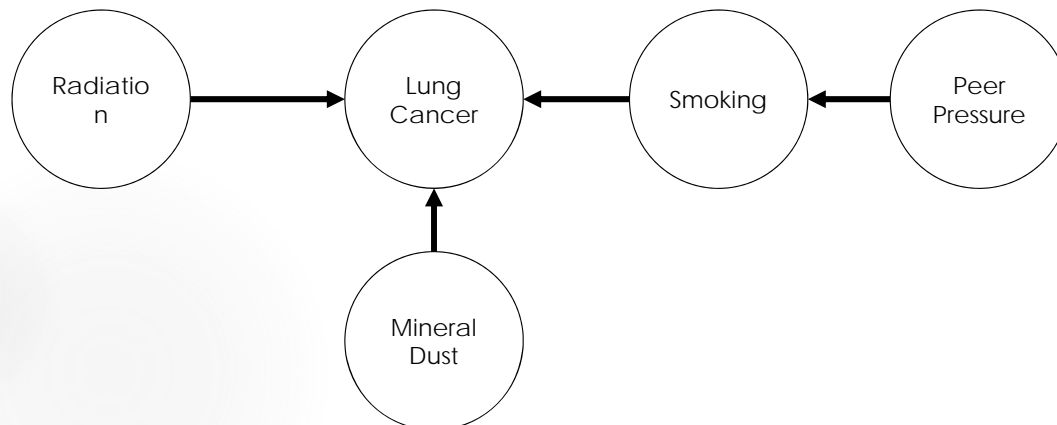


Research Methodology Workshop
22-24 April 2014



الجامعة الأردنية
Amman, Jordan

Example 2

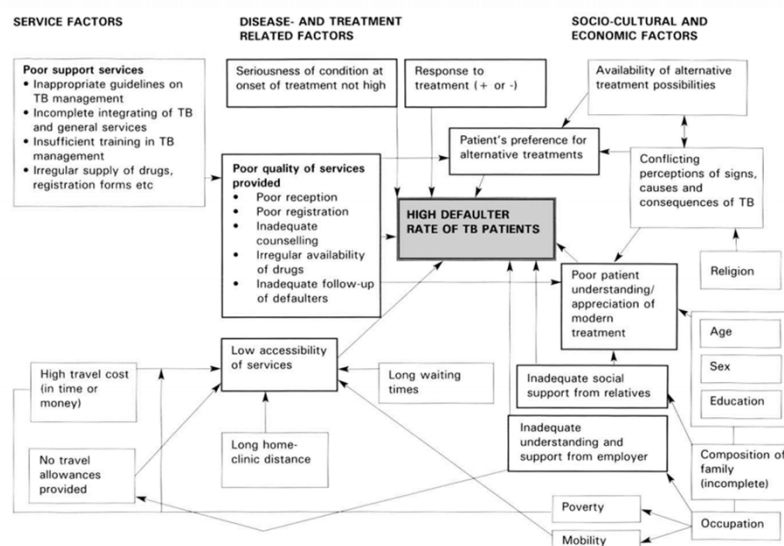


Research Methodology Workshop
22-24 April 2014



Ministry of Health
Government of Kenya

Figure 4.4: Revised problem analysis diagram of factors contributing to high defaulter rate among TB patients

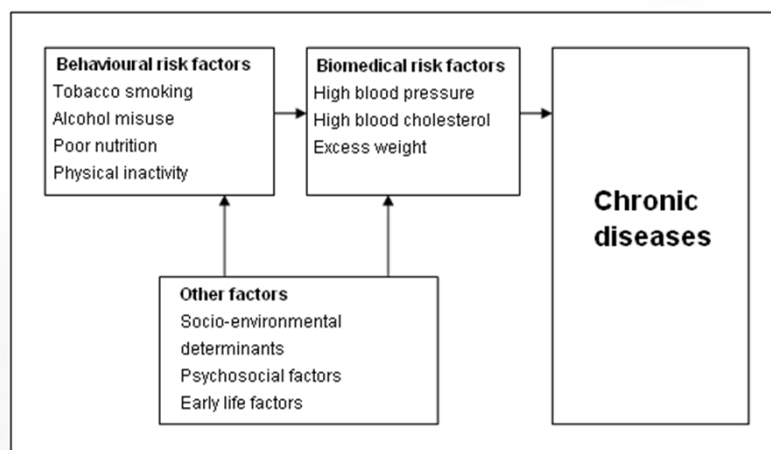


Research Methodology Workshop
22-24 April 2014

Varkevisser, C. M., Pathmanathan, I., & Brownlee, A. T. (2003). Designing and conducting health systems research projects: Kit Publishers.



Ministry of Health
Government of Kenya



المملكة العربية السعودية
Ministry of Health
Kingdom of Saudi Arabia

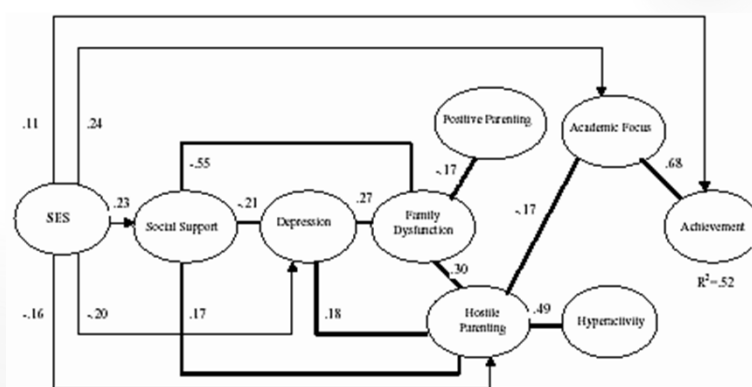


Figure 3 Structural Equation Model Showing the Relationship Between Family Processes, Child Characteristics, and Achievement for Girls Aged 6 to 11 years



المملكة العربية السعودية
Ministry of Health
Kingdom of Saudi Arabia

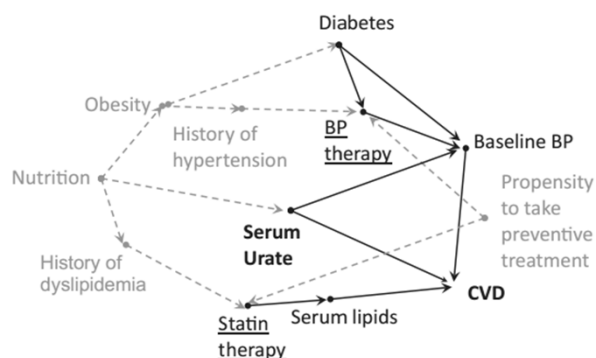


Figure 1: A directed acyclic graph (DAG), characterising the direction of selected influences* on risk of developing cardiovascular disease. BP- Blood pressure; CVD-Cardiovascular disease.

*Dark font variables are those which are observed, whereas grey font represent unobserved. Drug treatment variables, which we consider colliders, which introduce the influence of unobserved variables not directly linked with the outcome, are underlined.

Thornley S (2012)
Causation and Statistical Prediction:
Perfect Strangers or Bedfellows?
J Biom Biostat
3:e115.
doi:10.4172/2155-
6180.1000e115

Research Methodology Workshop
22-24 April 2014



Group Work #1 (2 hours)

1. Suggest the title of your research
2. State 3 reasons why your study is important
3. Identify the outcome, explanatory factors & the confounders (if applicable)
4. Build the conceptual framework
5. State the specific objectives of the study



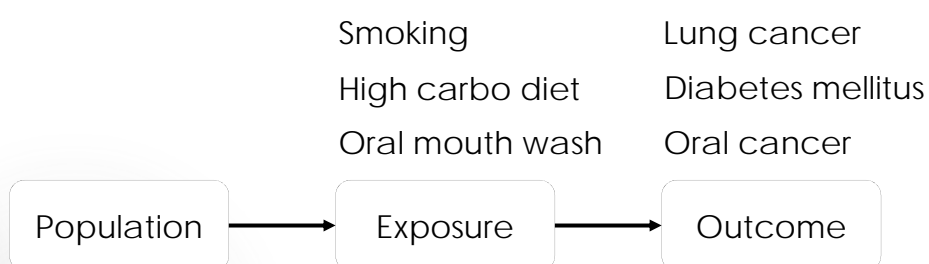
Research Methodology Workshop
22-24 April 2014

Part 3 – Study design



الجامعة الأردنية
The Jordanian University

Exposure & Outcome



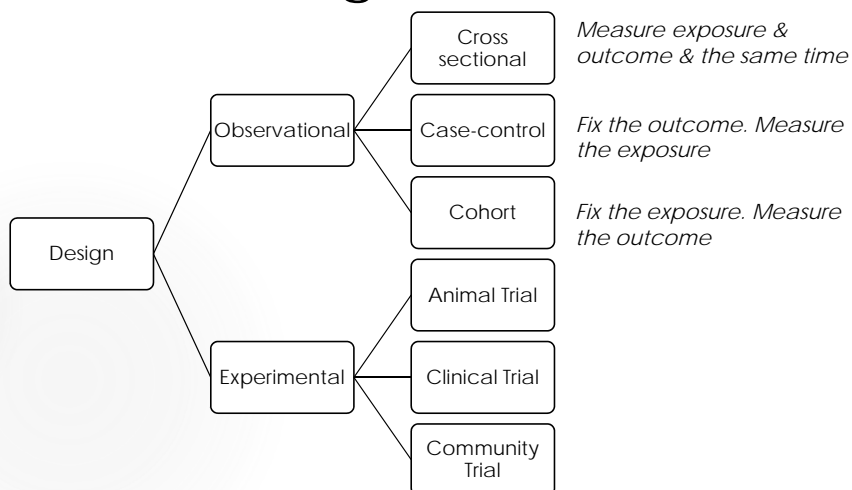
History → Future

To proof causation, exposure must precedes outcome



الجامعة الأردنية
The Jordanian University

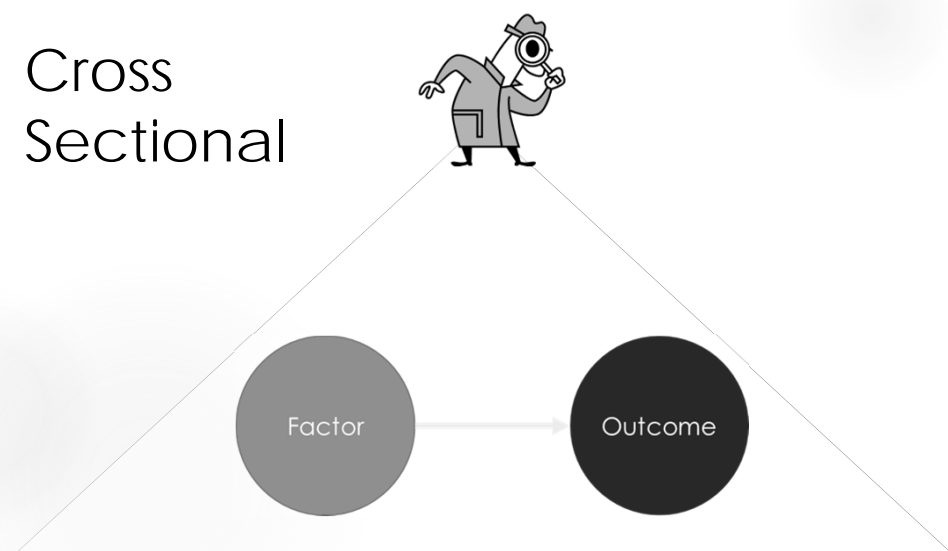
Research design



Research Methodology Workshop
22-24 April 2014



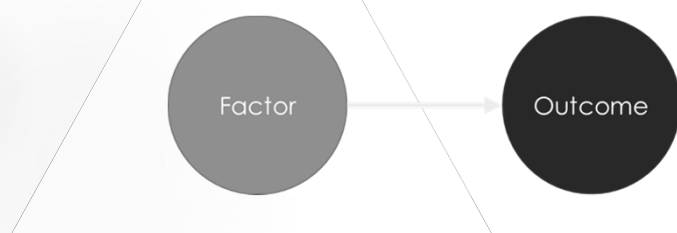
Cross Sectional



Research Methodology Workshop
22-24 April 2014



Case Control



الجامعة الإسلامية
جامعة القدس

Research Methodology Workshop
22-24 April 2014

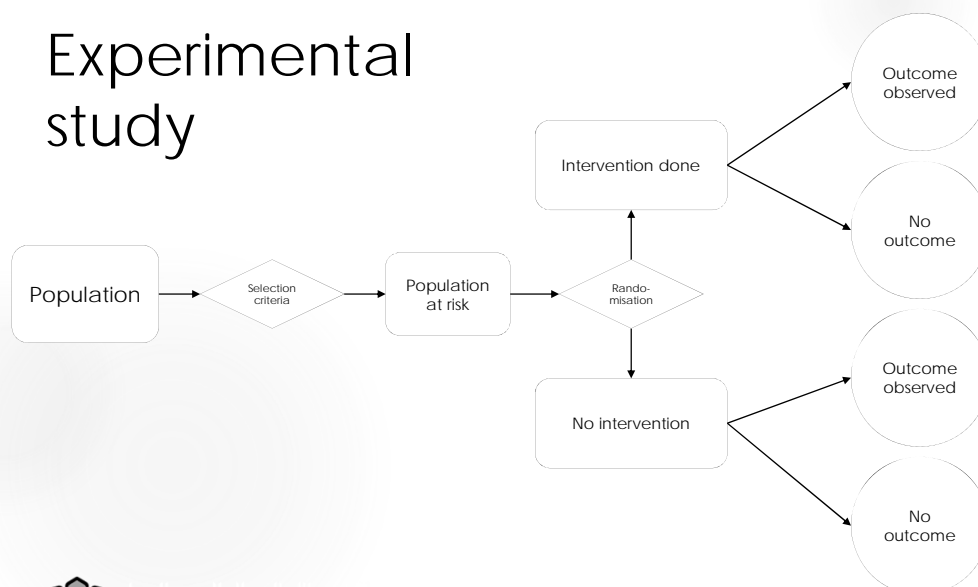
Cohort



الجامعة الإسلامية
جامعة القدس

Research Methodology Workshop
22-24 April 2014

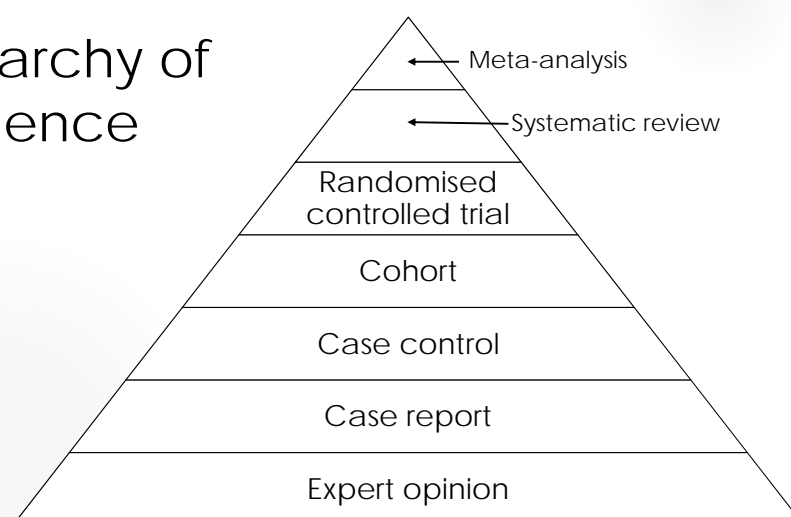
Experimental study



Research Methodology Workshop
22-24 April 2014

Research Methodology Workshop
22-24 April 2014

Hierarchy of Evidence



Research Methodology Workshop
22-24 April 2014

Research Methodology Workshop
22-24 April 2014

Cross sectional study

- ▶ Population based – represent population
- ▶ Measure exposure & outcomes at the same point in time – No temporal association
- ▶ Impossible to infer causality
- ▶ Prevalence study – measure magnitude or burden of disease
- ▶ Descriptive
- ▶ Repeated cross-sectional study ~ pseudo-longitudinal e.g. British Association for the Study of Community Dentistry (BASCD) guidance on sampling for surveys of child dental health. A BASCD coordinated dental epidemiology programme quality standard (Pine et al. 1997)



British Association for the Study of Community Dentistry (BASCD)

Research Methodology Workshop
22-24 April 2014

Cross sectional study

Advantages

- ▶ Measure prevalence of a population
- ▶ Measure multiple exposures & outcomes
- ▶ Relatively inexpensive
- ▶ Relative shorter time

Disadvantages

- ▶ No temporal association – no inference to causality
- ▶ Prevalence-incidence bias (Nyman bias) e.g. if smokers die due to AMI faster, a cross-sectional study will reveal less smoker among AMI patients
- ▶ Health workers effect e.g. when survey done from house to house, only health respondent are available in their home/office



British Association for the Study of Community Dentistry (BASCD)

Research Methodology Workshop
22-24 April 2014

Cross sectional study - Example

- ▶ NHMS ~ Household study, all Malaysian (N=47,610 for 2006)
- ▶ NOHSA ~ Adult (>15) (N=14,444 for 2010)
- ▶ NMCS ~ GP vs. PHC (N~12,000)
- ▶ NHANES (US) - <http://www.cdc.gov/nchs/nhanes.htm>



الوزارة
الصحة
الماليزيا

Research Methodology Workshop
22-24 April 2014

Case control study

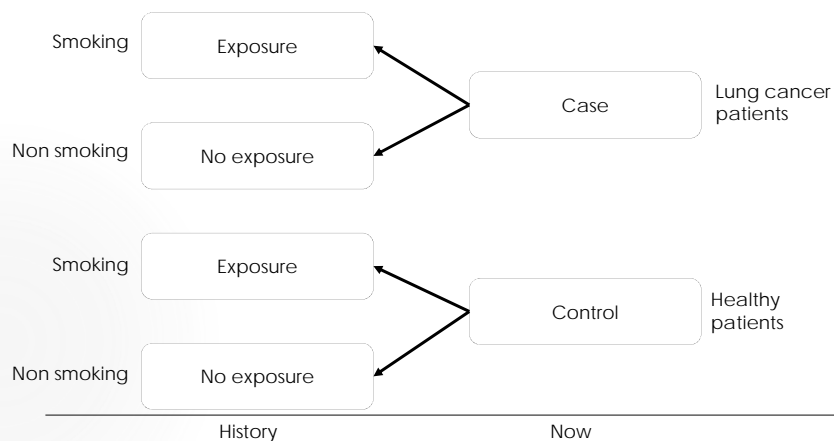
- ▶ Fix the outcomes, measure the exposures
- ▶ Longitudinal
- ▶ Retrospective
- ▶ Case = outcome of interest
- ▶ Control = comparing outcome



الوزارة
الصحة
الماليزيا

Research Methodology Workshop
22-24 April 2014

Case control study



22-24 April 2014



Research Methodology Workshop

	Lung Cancer	No Lung Cancer
Smoking	20 (18.2%)	90 (81.8%)
Not Smoking	5 (4.5%)	105 (95.5%)

$$\chi^2 (df=1) = 10.150, p = 0.001, OR = 4.7 (CI 95\% 1.7 - 13.0)$$

Because $p < 0.05$, we reject H_0 . Therefore there is a difference between smoker & non smoker



الجامعة الأردنية
البحر العلمي
البحر العلمي

Research Methodology Workshop
22-24 April 2014

Control

- ▶ Matched vs. Unmatched
- ▶ Matching ~ controls resemble the cases *with regard to certain characteristics (age, gender, SES etc)*
- ▶ Individual vs. Group matching
- ▶ Source – institution vs. population
- ▶ Ratio to cases ~ up to 4:1



البحر المتوسط
MEDITERRANEAN
Research Methodology Workshop

Case control study

Advantages

- ▶ Good for rare conditions or diseases
- ▶ Less time needed to conduct the study because the condition or disease has already occurred
- ▶ Measure multiple risk factors
- ▶ Can establish an association

Disadvantages

- ▶ Recall bias
- ▶ Not good for evaluating diagnostic tests because it's already clear that the cases have the condition and the controls do not
- ▶ It can be difficult to find a suitable control group



Example

- Tobacco use as a risk factor of myocardial infarction in 52 countries in the INTERHEART study: a case-control study (Teo 2006)

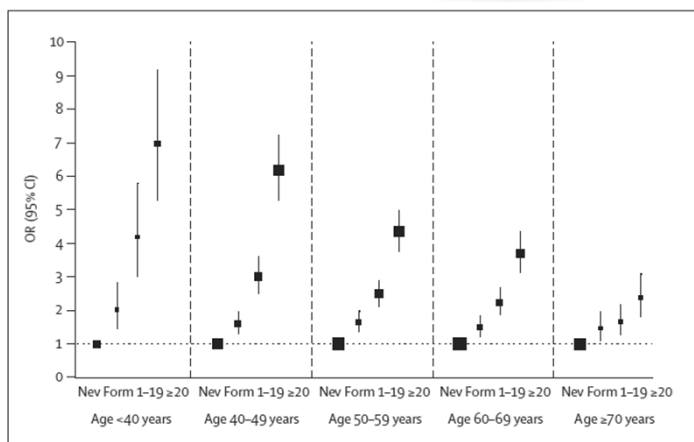


Figure 2: Risk of AMI associated with numbers smoked, by age group
p for interaction < 0.0001. Nev=never smokers. Form=former smokers. 1-19=currently smoking 1-19 cigarettes per day. ≥20=currently smoking 20 or more cigarettes per day.



Research Methodology Workshop

Research Methodology Workshop
22-24 April 2014

Cohort study

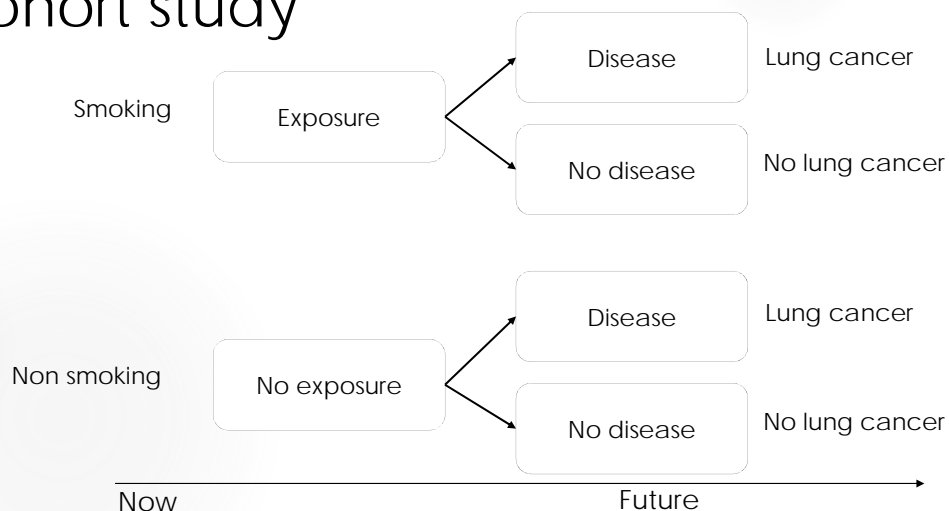
- Measure outcomes
- Compare incidence of a disease (or condition) among exposed and unexposed individuals over time
- Disease free at the onset (or inception)
- Repeated measurements ~ follow up
- Prospective vs. retrospective cohort



Research Methodology Workshop

Research Methodology Workshop
22-24 April 2014

Cohort study



Research Methodology Workshop
22-24 April 2014



Define cohort

- ▶ Both exposed & not exposed groups have equal chance to:
 - ▶ Develop disease
 - ▶ Be followed-up
- ▶ Types:
 - ▶ Representative – low exposed subjects
 - ▶ Enriched – high exposed subjects
 - ▶ Specific group – occupational, institution etc

Research Methodology Workshop
22-24 April 2014



Measurements

- ▶ Exposure
 - ▶ Carefully defined in advance
 - ▶ Standard measurement for both E+ & E- groups
- ▶ Outcome
 - ▶ Primary vs. Secondary outcome



المملكة العربية السعودية
Ministry of Health

Research Methodology Workshop
22-24 April 2014

Follow-up

- ▶ Keep participation at > 90%
- ▶ Equal likelihood to detect disease in all subjects
- ▶ Active vs. Passive follow-up
- ▶ Blinding



المملكة العربية السعودية
Ministry of Health

Research Methodology Workshop
22-24 April 2014

Example

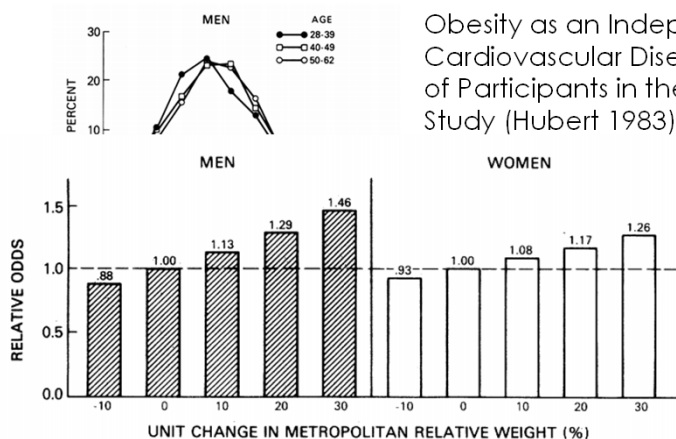


FIGURE 10. The relative odds of developing cardiovascular disease corresponding to degrees of change in Metropolitan Relative Weight between age 25 years and entry into the Framingham Study. The odds ratios reflect adjustments for the effects of relative weight at age 25 years and age and risk factor levels at exam 1.

Research Methodology Workshop
22-24 April 2014



Cohort study

Advantages

- ▶ Infer causality
- ▶ Measure multiple outcomes
- ▶ Study rare exposure
- ▶ Measure incidence

Disadvantages

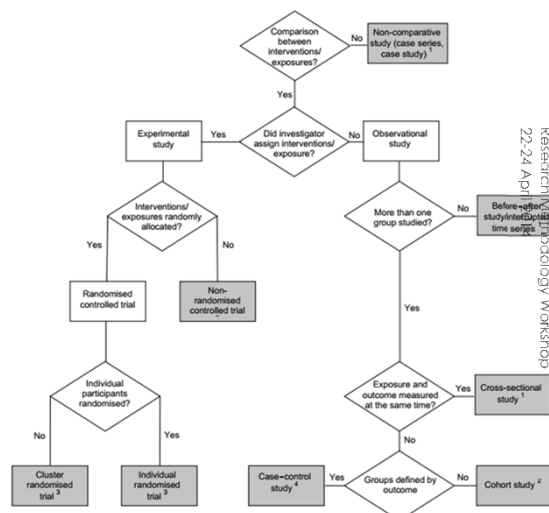
- ▶ Costly
- ▶ Loss to follow up
- ▶ Large sample size for rare outcomes
- ▶ Selection bias

Research Methodology Workshop
22-24 April 2014



TABLE 2.3 CHOICE OF STRATEGY

Basis	Cohort	Case-control	Cross-sectional
Rare condition	Not practical	Bias	Not appropriate
To determine a precise risk	Best	Only estimate possible	Gives relative prevalence, not incidence
To determine whether exposure preceded disease	Best	Not appropriate	Not appropriate
For administrative purposes	Not appropriate	Not appropriate	Best
If attrition is a serious problem	Not appropriate	Attrition is usually minimal	Attrition may have occurred before the study
If selective survival is problem	Best	Not appropriate	Not appropriate
If all factors are not known	Best	Not appropriate	Less appropriate
Time and money	Most expensive	Least expensive	In between



Part 4 – Sample & sampling method



Sampling & sample size

- ▶ Type of sampling – *Random vs. Non random*
- ▶ Sample size – *Based on objective & research design*

Research Methodology Workshop
22-24 April 2014



الجامعة الإسلامية
جامعة القدس

Before we sample

Determine study place, duration & subjects

- ▶ Describe study place – especially if plan to represent a population
- ▶ State time & duration
- ▶ Who or what are the subjects – population, people, animal etc.

Research Methodology Workshop
22-24 April 2014



الجامعة الإسلامية
جامعة القدس

Subjects

- ▶ Target population
- ▶ Study population
- ▶ Sampling frame
- ▶ Sampling unit
- ▶ Observation unit



السلطنة
سلطنة
سلطنة

Research Methodology Workshop
22-24 April 2014

Example – NHMS III 2006

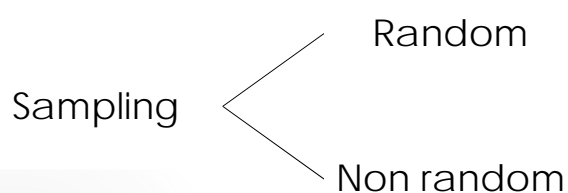
- | | |
|---------------------|--|
| ▶ Target population | <i>All Malaysian</i> |
| ▶ Study population | <i>Household up to strata 6</i> |
| ▶ Sampling frame | <i>List of Enumeration Block & Living Quarters</i> |
| ▶ Sampling unit | <i>Enumeration Block & Living Quarters</i> |
| ▶ Observation unit | <i>All household in the selected Living Quarters</i> |



السلطنة
سلطنة
سلطنة

Research Methodology Workshop
22-24 April 2014

Sampling method



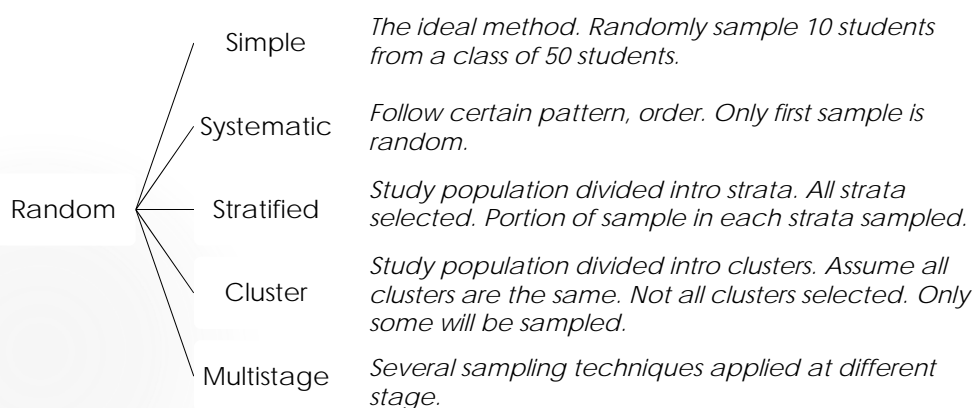
Random – equal chance to be selected

Research Methodology Workshop
22-24 April 2014



الجامعة الإسلامية
Islamic University

Random sampling

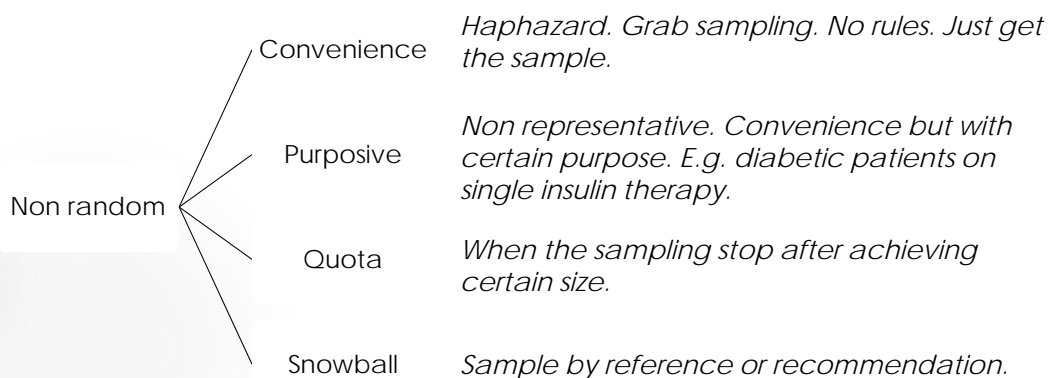


Research Methodology Workshop
22-24 April 2014



الجامعة الإسلامية
Islamic University

Non random sampling



Research Methodology Workshop
22-24 April 2014



Ministry of Health
Malaysia

Example – NHMS III 2006

Two stage stratified random sampling

- ▶ Target population *All Malaysian*
- ▶ Study population *Household up to strata 6*
- ▶ Strata *State & location (urban or rural)*
- ▶ Clusters *Enumeration Block & Living Quarters*
- ▶ Sampling frame *List of Enumeration Block & Living Quarters*
- ▶ Sampling unit *Enumeration Block & Living Quarters*
- ▶ Observation unit *All household in the selected Living Quarters*
- ▶ Sample distribution *Proportionate to size*

Research Methodology Workshop
22-24 April 2014



Ministry of Health
Malaysia

Sample size

- ▶ How many sample required
- ▶ An estimate
- ▶ Adequate size to
 1. Represent population
 2. Test hypothesis
- ▶ Size affects duration & budget of the research
- ▶ Do not sample more than required



Sample size

- ▶ Sample size depends on
 1. Objective of the study
 2. Study design (design effect)
 3. Sampling method
 4. Expected (& precision) effect size
 5. Variability of sample
 6. Non-response rate
- ▶ Even the expected outcomes/effect size are estimates



Sample size – formula

- ▶ No single formula for all
- ▶ Depending on purpose – single proportion, compare two proportions, compare 3 proportions etc.
- ▶ Use of software or calculators e.g. PS Power and Sample Size, PASS etc.
- ▶ Anticipate non response, drop-out, loss to follow up, death (esp. animal study)



Research Methodology Workshop

22-24 April 2014

Example – Single proportion

- ▶ $N = \frac{z^2 p(1-p)}{d^2}$, where
- ▶ N is the sample size,
- ▶ z = z value for intended confidence interval,
- ▶ p is the estimated proportion (in decimal) &
- ▶ d is the precision of p (deviation from p) (in decimal)



Research Methodology Workshop

Research Methodology Workshop
22-24 April 2014

Example – Single proportion

- ▶ Calculate sample size if you wish to do a study estimating the prevalence of DM as 20% with 5% variation at 95%CI (z for 95%CI is 1.96) anticipating 20% non response
- ▶ Answer: 246, ~ 250 then add 20% ~ 300 samples required



المملكة العربية
السعودية
Ministry of Health
Kingdom of Saudi Arabia

Research Methodology Workshop
22-24 April 2014

Group Work #2

1. What is the best study design? Give one reason
2. Describe the sample planned for your study – target, study, sampling frame, sampling unit & observation unit
3. Calculate the sample size required



المملكة العربية
السعودية
Ministry of Health
Kingdom of Saudi Arabia

Research Methodology Workshop
22-24 April 2014

Part 5 – Data collection



UNIVERSITY OF THE
WESTERN CAPE

Plan for data collection

- ▶ Identify all variables of interest – *produce detail data dictionary – declare all definitions & the measure types*
- ▶ Must check **validity & reliability** of research tools – including questionnaires
- ▶ Training – ensure similar ways of collecting data
- ▶ Key word here - STANDARD



UNIVERSITY OF THE
WESTERN CAPE

Research instruments

- ▶ Must be both valid and reliable
- ▶ Valid – content, face, criterion, construct etc
- ▶ Reliable - repeatability
- ▶ Must declare all tools used
- ▶ Including the questionnaire



Questionnaire or Record Form

- ▶ Face-to-face vs. self-filled (online or paper-based) vs. postal vs. telephone
- ▶ Specific questions (fulfill the objectives)
- ▶ Open vs. closed ended
- ▶ Structured usually closed-ended – e.g. dichotomous, Likert scale, multiple choice.
- ▶ Response code – e.g. 1=Male, 2=Female



MODUL R
MODULE R

KENCING MANIS
DIABETES

Butiran & Tajuk / Questions & Titles

Kategori Kod
Coding Category

Tajuk
Title

1. Adakah pengamal pesakit diabetes yang pernah mengambil ubat diabetes? / Have you ever taken any medication for diabetes?

2. Adakah pesakit diabetes yang sedang mengambil ubat diabetes? / Have you ever taken any medication for diabetes?

3. Dalam masa 4 minggu ini, adakah anda masih mengambil ubat diabetes yang sama? / In the last 4 weeks, were you still taking any medication for diabetes?

4. Apakah jenis ubat yang anda ambil? / What are the medications that you are currently taking?

5. Apakah jenis ubat insulin yang anda ambil? / What are the insulin medications you are currently taking?

6. Adakah anda pernah dimasukkan dalam kalangan pesakit yang telah dimasukkan ke hospital? / Have you been admitted by a medical personnel to do anything?

NOHSA 2010 SURVEY FORMAT
(FRONT PAGE OF FOLDER)

1. DATE OF SURVEY

2. IDENTIFICATION STATUS

3. STATE

4. ADMINISTRATIVE DISTRICT (DP)

5. CENSUS DISTRICT (DB)

6. ENUMERATION BLOCK (BP)

7. LIVING QUARTER (TK)

8. ADDRESS OF LG (TK)

9. STRATA

10. HOUSEHOLD (BH)

11. MEMBER NO

GENERAL INFORMATION

12. DATE OF BIRTH

13. AGE (AS ON LAST BIRTH)

14. GENDER

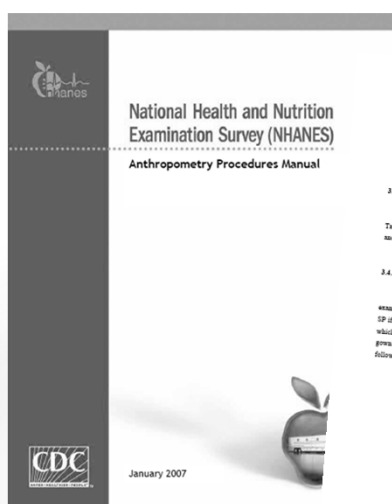
15. ETHNIC GROUP

16. HIGHEST EDUCATIONAL LEVEL ATTAINED

17. MONTHLY INDIVIDUAL INCOME



الدراسة الوطنية
للصحة والتغذية
National Health and Nutrition
Examination Survey (NHANES)



6. Record all measurements (except children < 2 years). Record children measurements in mm. For quality control purposes, the ring fall below the 1st percentile or above the 95th PERCENTILE will not be used.

2.4 Examiners Procedures:

This section describes the anthropometry examination procedure. Table 2-1, the protocol includes a total of 10 body measures. Some are and a minimum of eight measures, depending on the age of the participant.

2.4.1 Weight:

Participants will be weighed in kilograms using a digital scale. The scale will display the weight in both kilograms and pounds. If the scale is not working, the examiner should use a manual scale. Participants should wear the same clothes, shoes, and socks. The procedure for obtaining a weight is:

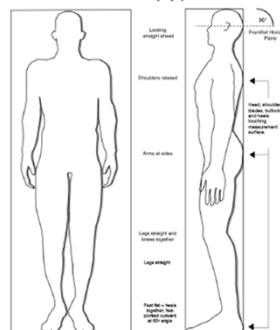
1. Position the SP: After the examinee fully enters the scale, the SP should be in the center of the scale.
2. Capture the result: After the SP is correctly positioned, the examiner should capture the result on the scale.

In addition, exam staff should follow the procedures described in:

- Small children: Infants and toddlers who cannot be weighed with the stadiometer of an adult. Either a baby scale or a toddler scale should be used. The child is then weighed on the scale. In this way, the scale will read only the child's weight. The

In the event of a power outage or if the stadiometer malfunctions, take the headpiece to the top of the measurement column and place the height using the tape measure secured on the right side of the measurement column. Call the result to the recorder, who will manually enter this number in the appropriate DSIS field.

Exhibit 2-2. Standing height position



3-9



الدراسة الوطنية
للصحة والتغذية
National Health and Nutrition
Examination Survey (NHANES)

Data dictionary

- ▶ List of variables
- ▶ Definitions & working definitions
e.g. When you say a subject is diabetes? Hypertensive?
- ▶ Cut-of-point e.g. How many age category
- ▶ Outcome vs. factors
- ▶ Type of measures – dependent or independent
- ▶ Coding e.g. 1=Male, 2=Female



Data dictionary - Example

<p>ANALYTIC AND REPORTING GUIDELINES: The Third National Health and Nutrition Examination Survey, NHANES III (1988-94)</p> <p>October, 1996</p> <p>National Center for Health Statistics Centers for Disease Control and Prevention Hyattsville, Maryland</p>	<p>SECTION I Key variables for analysis</p> <p>The categories and descriptions for the following 4 variables are consistent with the survey design and should be used in analysis, publication, and presentation of the NHANES III data. These descriptions are consistent with the 'Standardized and Suggested Reporting Categories' that were reported in the Third Report on Nutrition Monitoring (United States CDC, 1993). These categories and descriptions may be collapsed further for selected analyses, except for those or more variables are used simultaneously. The categories defined in this section should be arranged in the order in which they are used simultaneously. For example, there are 4 categories defined in this section for age, sex, race, and Hispanic-American status and family size. For all age variables, any findings with confidence. Thus, this age variable is consistent with the age group 10-19 years, or both sexes present with the age group 10-19 years, or both sexes present with the age group 10-19 years, or both sexes present with the age group 10-19 years. Any exceptions to these guidelines should be noted. The following list includes the labels and SAS variable names from the NHANES III documentation.</p> <p>Age: NHANES and NHANES (Age at interview)</p> <p>Total 2-11 months (DO NOT use < 1 year which includes infants months)</p> <p>1-11 years 2-11 years 3-11 years 4-11 years 5-11 years 6-11 years 7-11 years 8-11 years 9-11 years 10-11 years 11-11 years 12-11 years 13-11 years 14-11 years 15-11 years 16-11 years 17-11 years 18-11 years 19-11 years 20-11 years 21-11 years 22-11 years 23-11 years 24-11 years 25-11 years 26-11 years 27-11 years 28-11 years 29-11 years 30-11 years 31-11 years 32-11 years 33-11 years 34-11 years 35-11 years 36-11 years 37-11 years 38-11 years 39-11 years 40-11 years 41-11 years 42-11 years 43-11 years 44-11 years 45-11 years 46-11 years 47-11 years 48-11 years 49-11 years 50-11 years 51-11 years 52-11 years 53-11 years 54-11 years 55-11 years 56-11 years 57-11 years 58-11 years 59-11 years 60-11 years 61-11 years 62-11 years 63-11 years 64-11 years 65-11 years 66-11 years 67-11 years 68-11 years 69-11 years 70-11 years 71-11 years 72-11 years 73-11 years 74-11 years 75-11 years 76-11 years 77-11 years 78-11 years 79-11 years 80-11 years 81-11 years 82-11 years 83-11 years 84-11 years 85-11 years 86-11 years 87-11 years 88-11 years 89-11 years 90-11 years 91-11 years 92-11 years 93-11 years 94-11 years 95-11 years 96-11 years 97-11 years 98-11 years 99-11 years 100-11 years</p>	<p>Poverty Index (Poverty income ratio, PIR): INHPIR</p> <p>This is a calculated variable based on family income and family size using tables published each year by the Bureau of the Census in a series "Current Population Reports" on poverty in the United States. This is the best income variable to use when comparing data over time because it is "relatively" standardized for inflation and other factors. However, the method of calculation has been changed slightly over time. The primary reporting categories are:</p> <p>0.000-0.999 (below poverty) 1.000 and above (at or above poverty)</p> <p>Again, there are a significant number of persons for whom this variable cannot be calculated.</p> <p>For some specific analyses, use of the Data Food Assistance Program (Special Supplemental Nutrition Program for Women, Infants, and Children (SIC), Food Stamp Program, School Lunch and Breakfast Program) eligibility cut points of 1.000 or 1.400 is acceptable. The categories to use in these options are:</p> <p>0.000-1.399 (low) 1.400-1.999 (middle) 2.000 and above (high)</p> <p>or</p> <p>0.000-1.999 (low) 2.000-2.999 (middle) 3.000 and above (high)</p> <p>Region: INHPIR (Census region)</p> <p>Northeast (code 1) Midwest (code 2) South (code 3) West (code 4)</p> <p>These four regions are defined by the Census and can be combined as needed in analyses. For example, sample size may be too small for Mexican Americans in the Northeast or the Midwest regions; therefore, codes 1 and 2 can be combined in some analyses.</p>
--	---	--



Suggested details to include

1	Name	The name normally required in computer such as in database and statistical analysis. Name can be in one short word e.g. agecat for Age Category
2	Label	The name that can appear in table, graph or report
3	Definition	The definition used in the research. It is advisable to include references used
4	Instrument used	When relevant, we can describe the instrument used which include the brand and the method of calibration if relevant
5	Level of measurement	Should specify either it is nominal, ordinal or continuous
6	Category option and code	If the variable is categorical, the options should be specified e.g. Gender; Male=1, Female=2
7	Unit of measurement	If the variable is numerical, we should specify its unit e.g. mmol/L, mg/dL
8	Precision of measurement	How precise the variable is measured e.g. Age is measured to the nearest 1 year old. Income is measured to the nearest RM100
9	Data linkage	If this variable is related to other variable, we can specify here e.g. Missing value (Question on pregnancy) if respondent is Male (Question on Gender)



Ministry of Health
Malaysia

Research Methodology Workshop
22-24 April 2014

Data dictionary is
very important!!



Ministry of Health
Malaysia

Research Methodology Workshop
22-24 April 2014

Part 6 – Statistical analysis plan



UNIVERSITY OF THE
WESTERN CAPE

Statistical analysis plan

- ▶ Based on objective, especially the specific objectives – orderly manner
- ▶ Descriptive vs. analytical analysis
- ▶ The product is dummy table
- ▶ State all statistical tests planned to be used
- ▶ State significant level
- ▶ State software used
- ▶ Engage a statistician from the beginning!



UNIVERSITY OF THE
WESTERN CAPE

Dummy table - example

Objective

To compare blood glucose level between gender

Variables involved

Variable label	Working definition (linkage data)	Status	Variable name	Level of measurement	Category label (if relevant)	Variable Unit	Precision of measurement	Missing value
Blood glucose	As measured	Dependent	glu	Interval		mmol/L	0.1	999
Gender	As reported	Independent	sex	Nominal	1 = Male, 2 = Female			None

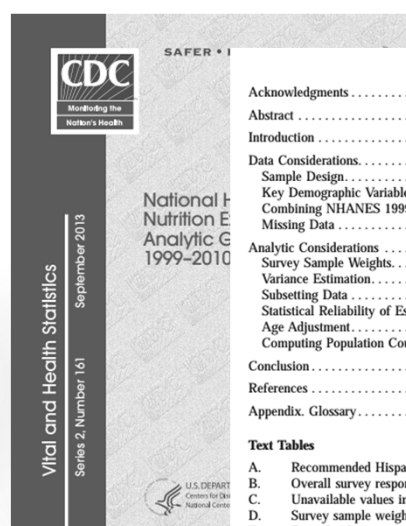
Statistical analysis

1. Check normality of *glu*
2. If *glu* Normal, run Independent sample t-test; if *glu* not Normal, run Mann-Whitney U-Test
3. Significance level = 0.05

Dummy table

	Mean (SD)	Statistics	df	P
Male	nn.n (n.n)	n.nnn	nn	0.nnn
Female	nn.n (n.n)			

SD = Standard deviation



Acknowledgments	iv
Abstract	1
Introduction	1
Data Considerations	1
Sample Design	1
Key Demographic Variables of Interest	3
Combining NHANES 1999-2010 Survey Cycles	5
Missing Data	5
Analytic Considerations	7
Survey Sample Weights	7
Variance Estimation	10
Subsetting Data	10
Statistical Reliability of Estimates	11
Age Adjustment	13
Computing Population Counts	13
Conclusion	14
References	14
Appendix. Glossary	15
Text Tables	
A. Recommended Hispanic subgroups for analyses: National Health and Nutrition Examination Survey, 1999-2010	2
B. Overall survey response rates for all ages: National Health and Nutrition Examination Survey, 1999-2010	6
C. Unavailable values in data: National Health and Nutrition Examination Survey, 1999-2010	7
D. Survey sample weights and their appropriate use: National Health and Nutrition Examination Survey, 1999-2010	8
E. Formulas for constructing weights: National Health and Nutrition Examination Survey, 1999-2010	9
F. Recommended sample sizes for analyses of complex survey data, by design effect and specified proportion	12



Data quality

- ▶ Valid value
e.g. age > 200 years, weight > 500 kg, pregnant male etc
- ▶ No missing value
- ▶ Relevant skip response
e.g. Not Applicable response for number of pregnancy for male respondent
- ▶ Declare method to ensure good data quality – e.g. double data entry



البحرین
National Institute for
Research in Reproductive Health

Research Methodology Workshop
22-24 April 2014

Group Work #3

- ▶ Prepare the data dictionary
- ▶ Describe what you will do with the data
- ▶ Prepare dummy table for each specific objective



البحرین
National Institute for
Research in Reproductive Health

Research Methodology Workshop
22-24 April 2014

Part 7 – Writing report



الجامعة الوطنية
UNIVERSITY OF THE PACIFIC

Before you write

- ▶ Decide the target audience
- ▶ Scientific publication or report
- ▶ Choose journal
- ▶ Study the format & requirement
- ▶ Separate text, table & graphics



الجامعة الوطنية
UNIVERSITY OF THE PACIFIC

The structure

Publication

- ▶ Title
- ▶ Abstract
- ▶ Keywords
- ▶ Introduction
- ▶ Method
- ▶ Results
- ▶ Discussion
- ▶ References



جامعة الباقية
University of the Pacific

Report/thesis*

- ▶ Title
- ▶ Abstract
- ▶ Introduction
- ▶ Literature review
- ▶ Objective
- ▶ Methodology
- ▶ Results
- ▶ Discussion
- ▶ References

* Institution specific

Research Methodology Workshop
22-24 April 2014

The suggested sequence

1. Based on specific objective, analyse the data & produce planned tables
2. Interpret & describe the results in Result section
3. Discuss in Discussion section
4. Answer the research questions
5. Complete the method & introduction
6. Finally, write the abstract



جامعة الباقية
University of the Pacific

Research Methodology Workshop
22-24 April 2014

Writing result

- ▶ Describe your result (no discussion)
- ▶ No reference (usually)
- ▶ Text vs. table vs. graphic (no redundancy)
- ▶ Text to summarise, Table for detail, Graphic to show trend
- ▶ May state relevant statistics done (if not mentioned in method)



Writing discussion

- ▶ Should answer the research questions mentioned in Introduction
- ▶ Discuss the result
- ▶ Do not repeat text as in Result
- ▶ May state limitation (but don't go overboard)
- ▶ Recommend
- ▶ Conclude



Part 8 – The administration



البحرین
البحرین

Other things to plan

1. Ethical consideration – consent form, advisory committee
2. Budget
3. Approval



البحرین
البحرین

In summary, what are the critical information

1. Specific objectives
2. Conceptual framework
3. Data dictionary
4. Dummy table & analytics guidelines



INSTITUT
NATIONAL DE STATISTIQUE
ET D'ÉTUDES ÉCONOMIQUES