

# Research Methods for MSc Students in Applied Human Nutrition

Tadesse Awoke (PhD)

Associate Professor

University of Gondar

Email: tawoke7@gmail.com

Course outline and schedule

# Contact Information

□ **Instructor:** Tadesse Awoke (Associate Professor)

Department of Epidemiology & Biostatistics

MSc. in Applied Statistics

MSc. in Biostatistics

PhD in Public Health

PhD in Biostatistics

□ **Contact:** Phone: +251910173308

Email: [tawoke7@gmail.com](mailto:tawoke7@gmail.com)

□ **Lectures:** Date and time: August 04---, 2018

Location: UoG

□ **Office hours:** Available upon request

# *Course Description*

- ❑ *The research methods course is designed to assist students in developing their thesis research, which they are required to undertake as part of the Masters Program.*
  - ❑ Students are expected to take the lead in identifying and pursuing the research area of their choice.
  - ❑ Each student shall identify a research area at the beginning of the course.
  - ❑ Choose researchable topic for their thesis work
  - ❑ Link the identified research problem with steps in developing research proposal
  - ❑ Help them to develop research proposal for the identified problem

# Course Objectives

- *By the end of the course students will acquire knowledge and skills of:*
  - Developing research proposal
  - Searching literature related to their research area
  - Data collection and data analysis
  - Managing research projects
  - Handling ethical issues
  - Writing thesis report.

# Research Methods Course Schedule Aug 04--, 2018

Date	Topics	Time
Day 1	Course Introduction	2:00-5:30
	Definition and types of research	
	Research topic selection criteria	
	Literature review	
	<b>Assignment</b>	
Day 1	Searching literatures from different sources	8:30-12:30
	Introduction to EndNote	
	Formulation of research objective	
	Choosing study design	
	<b>Assignment</b>	
Day 2	Data collection methods/techniques	2:00-5:30
	Questionnaire design	
	<b>Assignment</b>	

# Research Methods Course Schedule Aug 04--, 2018

Date	Topics	Time
Day 3	Choosing sampling technique & sample size calculation  <b>Assignment</b>	3:30-5:30
Day 4	Planning data analysis & choosing statistical techniques  Basic data management using SPSS  <b>Assignment</b>	8:00-12:30  4:30-5:30
Day 5	Research ethics  Writing a thesis proposal & Reviewing thesis proposal  Scientific paper writing & Reviewing thesis report  <b>EPI-INFO 7</b>  <b>SPSS/STATA</b>	8:00-12:30  2:00-4:00  4:00-5:30
-----	Proposal presentation and submission  <b>Assignment</b>	Students

# Methods of Instruction

- Lectures
- Case studies
- In-class practical exercises
- Group discussions.
- Individual Proposal development
- *Student presentation*

## Materials needed

- EndNote15
- EPI-INFO 7
- SPSS/STATA/SAS/R

## Course pre-requisite

-Epidemiology

-Biostatistics

## Attendance criteria

-Attendance of the lectures, group work, topic and proposal defense sessions is strictly compulsory

# Assessment criteria

- ❑ Class assignment-----(30%)
  - ❑ Research title selection -----5%
  - ❑ Introduction -----5%
  - ❑ Formulation of Objectives---5%
  - ❑ Methods -----10%
  - ❑ Ethics -----5%
- ❑ Proposal submission ----- (30%)
- ❑ Written exam-----40%
- ❑ Total -----100%

# References

- ❑ Kebede Y., Andargie G., Feleke, A., Awoke T. Module on Research Methods, University of Gondar, College of Medicine and Health Sciences Institute of Public Health, 2015
- ❑ Bowling A. Research Methods in Health. Investigating Health and Health Services. Open University Press, 2000
- ❑ John W. Creswell. Research Design. Qualitative, Quantitative, and Mixed Method Approaches (third edn). SAGE Publications, Inc., 2009
- ❑ Davies M. Brett. Doing a successful research project. Using Qualitative or Quantitative Methods. Palgrave macmillan, 2007
- ❑ Davies M. Brett. Doing a successful research project. Using Qualitative or Quantitative Methods. Palgrave macmillan, 2007
- ❑ Ann Aschengrau, George R. Seage III. Essentials of Epidemiology in Public Health (2<sup>nd</sup> edn). Jones and Bartlet Publishers, 2008
- ❑ Getu D., Tegbar Y., 2006: Research methodology for HSS

# Comments/Questions?

# Research Methods



Tadesse Awoke (PhD)

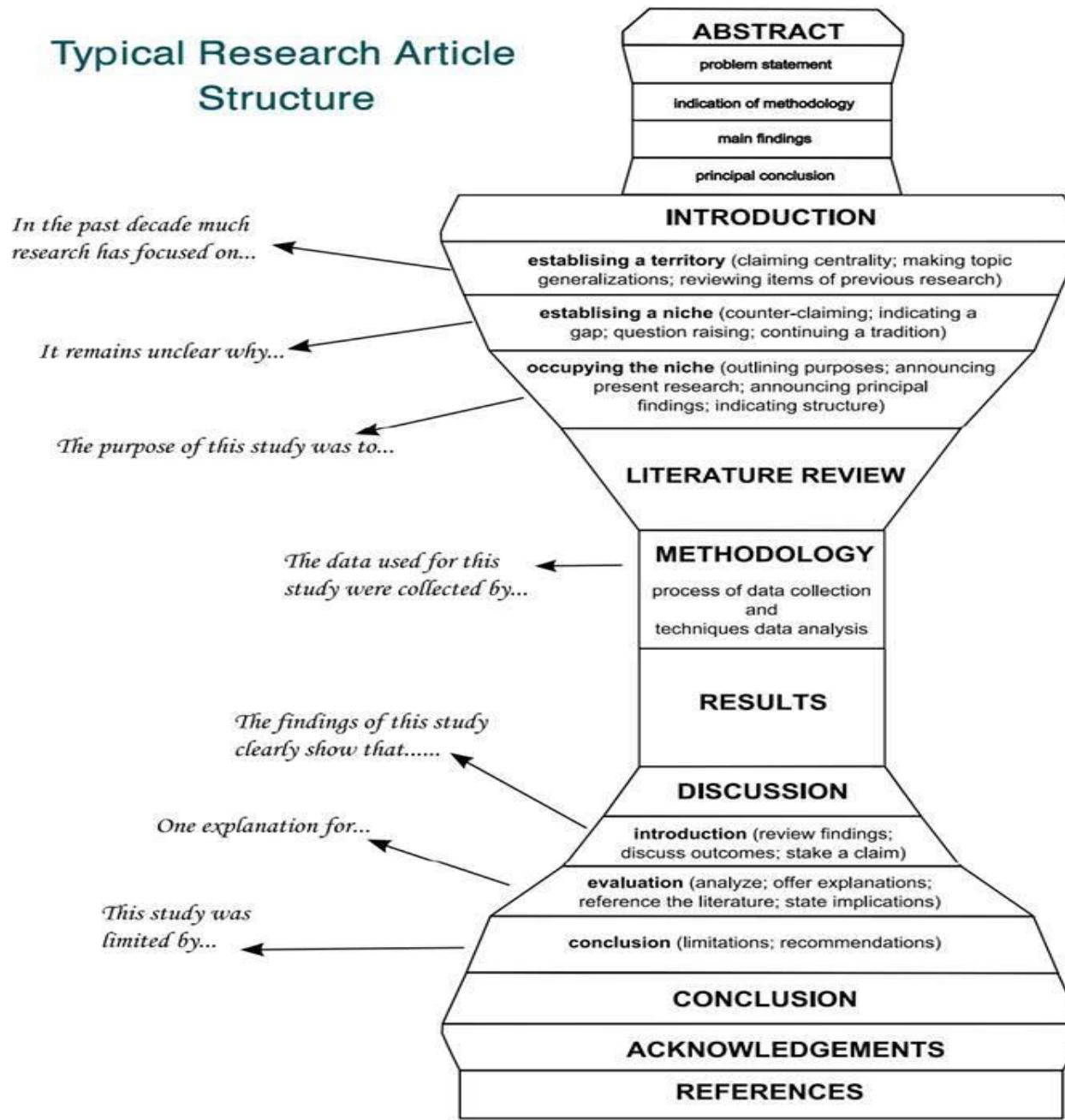
- Problem Identification
- Topic selection
- Types of research
- Components of research proposal
- Statement of the problem
- Literature review

# Learning Objectives

---

- At the end of this session the student will be able to:
  - Define health research, research design, and research methods
  - Understand the different types of research
  - Strategies underlying hypothesis formulation
  - Formulate research title
  - Develop statement and literature review
  - Develop justification for the proposed title of research

## Typical Research Article Structure



□ **Proposal 2017**  
pdf

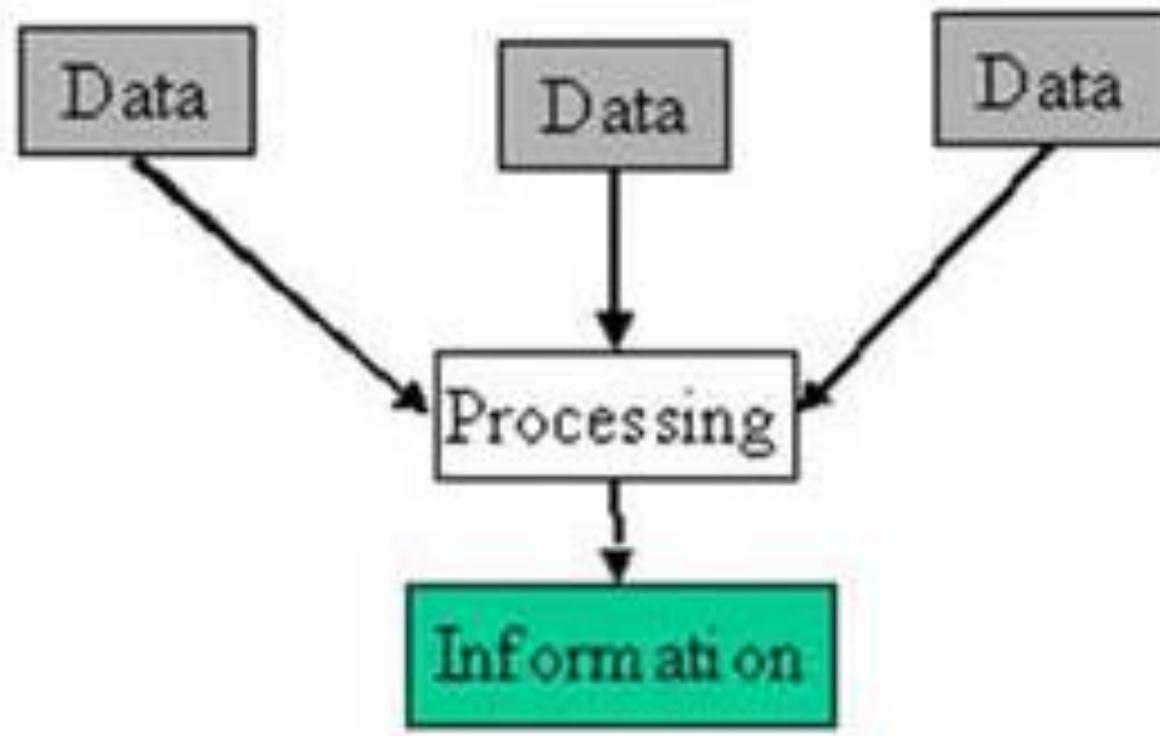
# Introduction (1)

---

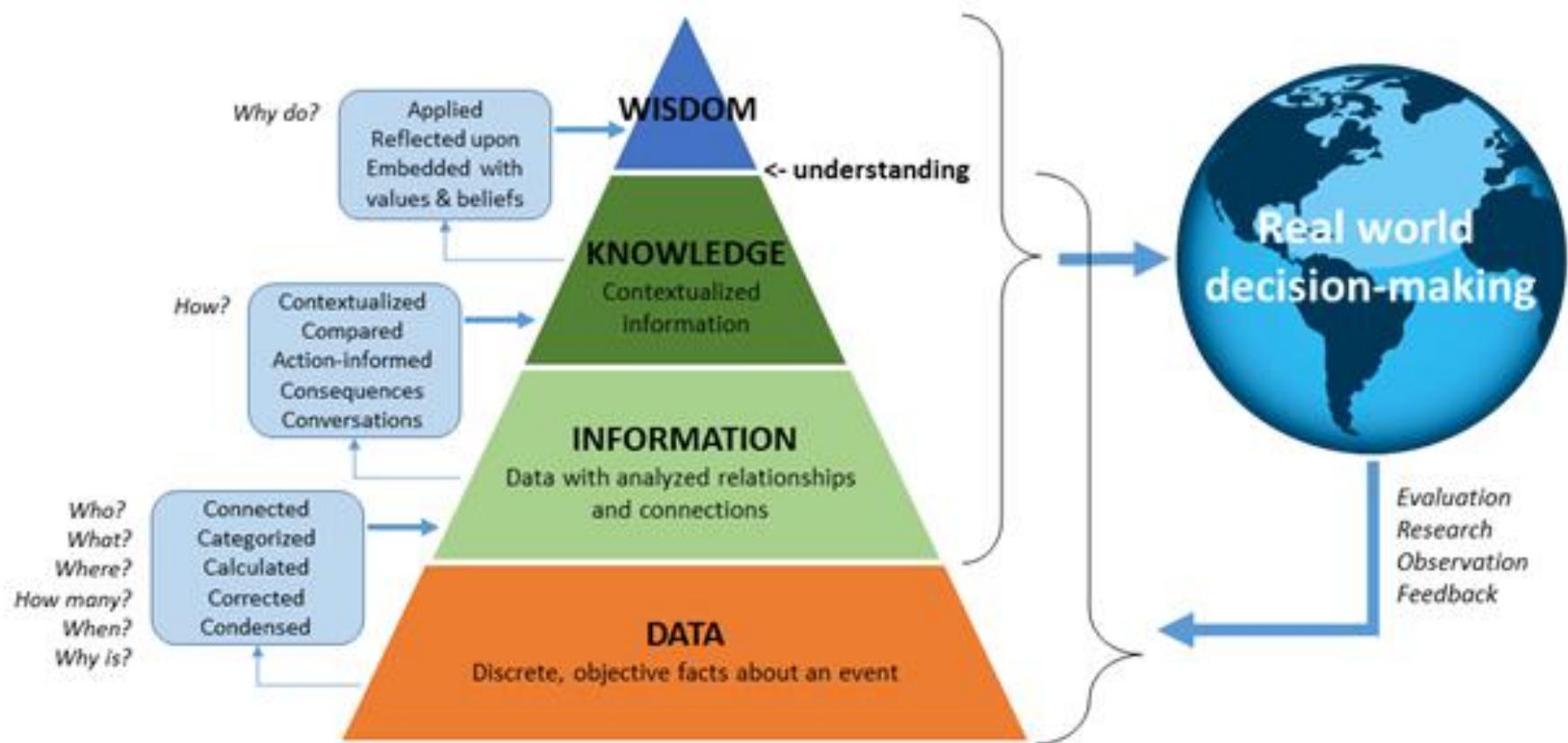
- **Research is defined as:**
  - **the systematic**
    - collection
    - analysis and
    - interpretation of data
- **to answer a certain question or solve a problem or add body of knowledge.**

# Data and Information

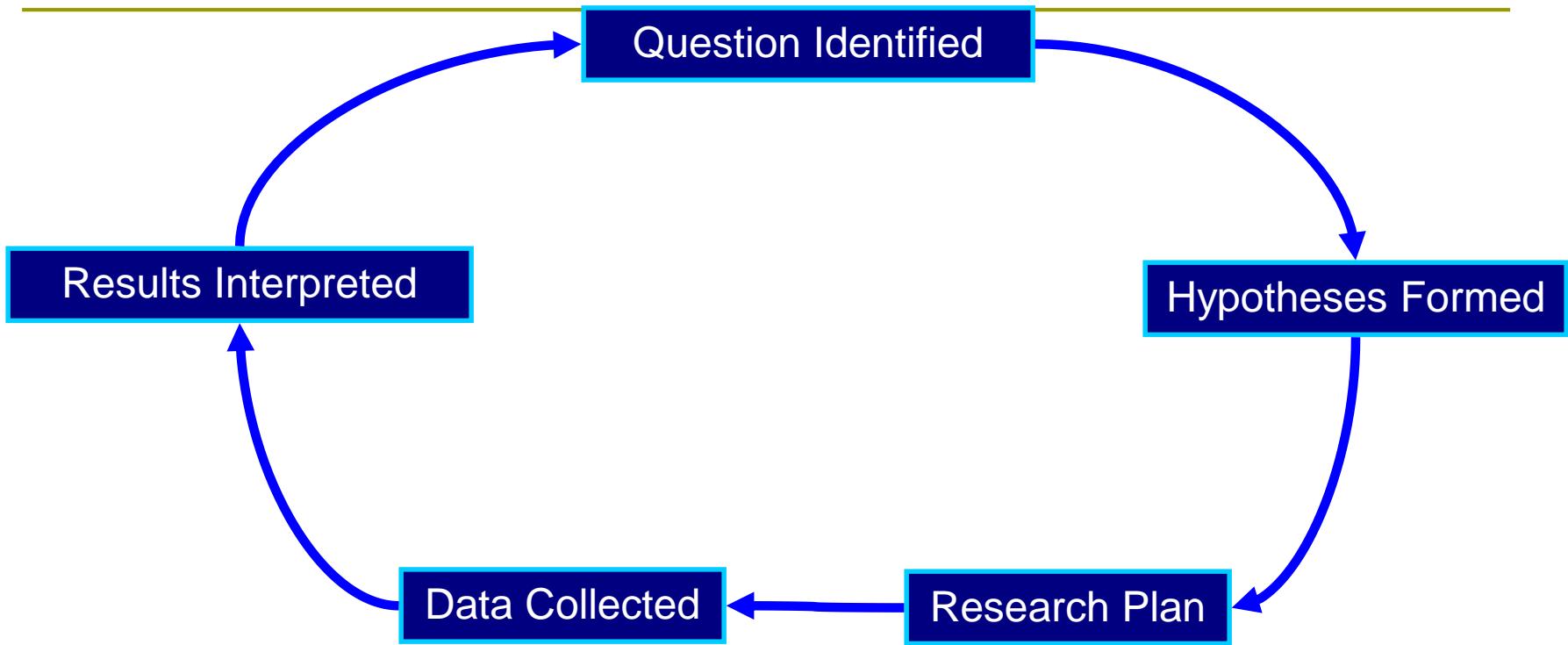
---



# Data and Wisdom



# Research Process



Closed-loop conceptualization of the research process (Drew, Hardman, and Hart, 1996)

# Health Systems Research (HSR)

---

- *Research that supports health development is known as Health Systems Research.*
- ***Health system Includes:***
  1. The individual, family and the community with its socio-cultural context
  2. Health care services
    - private and Public (governmental) sector
    - Health workers, health institutions, etc.
  3. Health related sectors : education, agriculture, etc.
  4. The international sector (*bilateral*) and multilateral donor agencies

# Types of Health Research

---

- Research can be classified in many different ways on the basis of the population, the knowledge it creates, the user group, the research problem it investigates, methodology of research and so on;
- **Based on population:**
  - Biomedical, clinical.....individual
  - Epidemiologic, Health Systems research.....population
- **Based on Objective:**
  - **Basic (Fundamental or traditional) research:** is necessary to generate new knowledge & technologies
  - **Applied (action) research:** to identify priority problems and to design and evaluate **Policies and programs.**

# Types of Health Research

---

- **Based on data type utilized:**
  - **Quantitative:** phenomena that can be expressed in terms of quantity
  - **Qualitative:** Qualitative phenomena (those relating to involving quality or kind)
- **Based on study design used and results obtained in the study:**
  - Descriptive research
  - Analytic research
- **Based on procedure**
  - Experimental
  - Nonexperimental

# Purpose of Health Research

---

- ❑ To generate knowledge essential to effectively promote the health of the population
- ❑ **Without that knowledge, effective action is impossible because it has no logical or empirical basis**

# Research must be

---

**Purposeful:** what do you want to be able to contribute?

**Targeted:** Who are the audiences?

**Credible:** consider sources information, method of data collection, personnel involved...

**Timely:** Is the information needed?

# Problem Identification

---

Where do we get research questions (topics)?

- **Observation (work place, others)**
- **Reading literatures**
- **Reports from health institutions**
- **Results of other researches**

# Problem Identification

---

## Techniques:

- **Select** an appropriate subject /problem for a research proposal that will be developed later
- 
- **Identify** criteria for selecting health-related problems to be given priority in research
- Use **a group consensus technique** to set priorities for research, applying the selected criteria on a number of research topics

# Criteria to researchable nature of a problem

---

1. **Presence of a discrepancy** between what exists and the ideal or planned situation
  
2. **Unclear reasons** for this difference.
  - It makes sense to develop research questions
  
3. **More than one possible answer** to a question or more than one solution to the problem.

# How do we choose a research topic?

---

- **Career development**
- **Priority research areas**
  - Relevance
  - Avoidance of duplication
  - Feasibility
  - Political acceptability
  - Applicability
  - Urgency of data needed
  - Ethical acceptability
- **Resource availability**

# Topic selection

---

## **Factors guiding the selection of a topic/title:**

- the possibility of the research
- the uniqueness of the research
- the scope of the research
- the profitability of the research
- the theoretical value of the research
- the practical value of the research

# Topic Formulation

---

- ❑ The research topic or title should be specific and clear.
- ❑ Answer the questions: What is my paper about? What techniques/ designs were used? Who/what is studied? What were the results?
- ❑ It is the focus of your research
- ❑ typically 10–12 words long

# Topic selection

---

- A good title is usually a compromise between conciseness and explicitness.
- Titles should be comprehensive enough to indicate the nature of the research.
- One good way to cut the length of titles is to avoid words that add nothing to a reader's understanding, such as "Studies on....," "Investigations....," or "Research on Some Problems in...."
-

# Titles Formulation Tips

---

- A good research paper title:
  - Condenses the paper's content in a few words
  - Captures the readers' attention
  - Differentiates the paper from other papers of the same subject area
- So here are three basic tips to keep in mind while writing a title:
  - 1] Keep it simple, brief and attractive
  - 2] Use appropriate descriptive words
  - 3] Avoid abbreviations and jargon

# Considerations in selecting a research problem

---

- Interest
- Magnitude
- Measurement of concepts
- Level of expertise
- Relevance
- Availability of data
- Ethical issues

# Examples

---

- ❑ Investigation on administrative System of Local NGOs, the case of Amhara region, Northwest Ethiopia, 2013.

# Assignment 1

---

- Think of three possible area of research which you feel that there is a problem
- Identify one of the problem and formulate the research topic that potentially is used in the thesis work
- Write the title of the research

# Overview of the Research Process



# Overview of research process

---

What is the problem?



Define your initial objective



Search literature



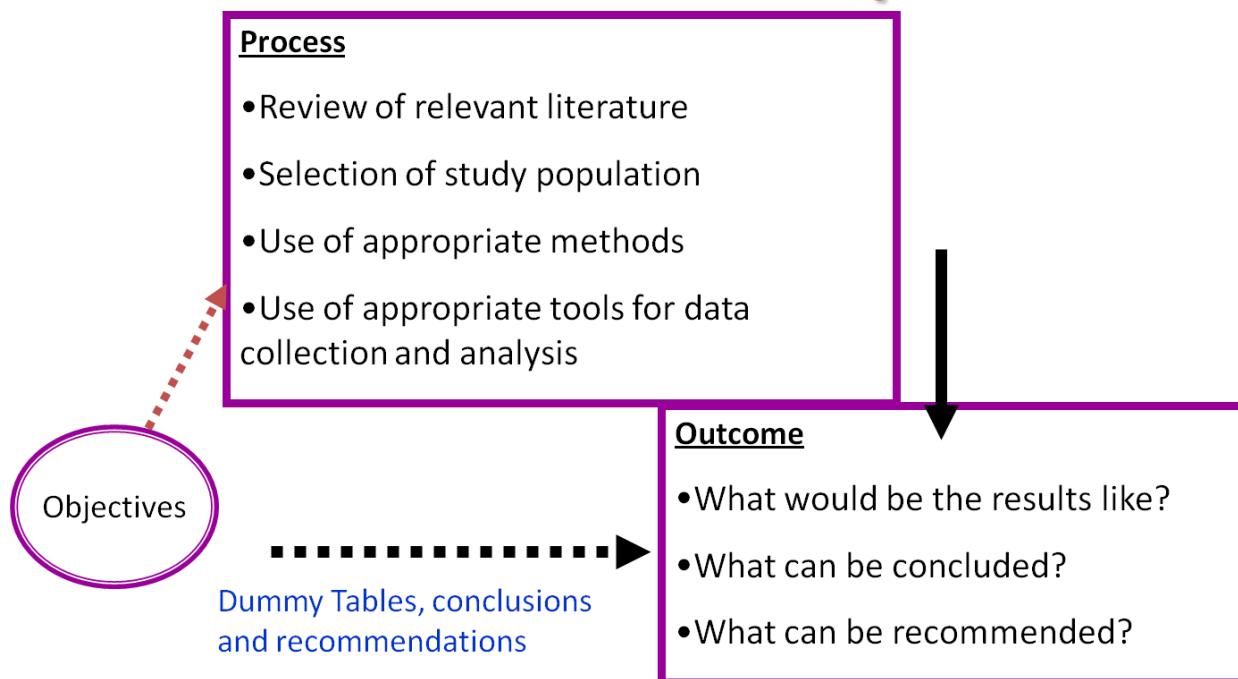
Modify your objective based on the information you get from literature



Write the full proposal guided by the objective

# Overview of research process

## Overview of research process



**Always focus on your objectives!!!**

# **Components of Research Proposal**

- |   |   |
|---|---|
| <b>1. Title</b>   | <b>8. Dissemination and Utilization of Results.</b> |
| <b>2. Summary</b>   | <b>9. Work plan</b>                                 |
| <b>3. Acronym</b>   | <b>10. Cost of the Project</b>                      |
| <b>4. Introduction</b> <ul style="list-style-type: none"><li>■ Statement of the Problem</li><li>■ Literature review</li><li>■ Justification</li></ul> | <b>11. References</b>                               |
| <b>5. Objectives</b>  | <b>12. Assurance of the investigator</b>            |
| <b>6. Methodology</b>   | <b>13. Advisor (approval)</b>                       |
| <b>7. Ethical Considerations</b>  | <b>14. Annex</b>                                    |

# Summary

---

Should reflect:

- ❑ Problem Statement
- ❑ Research objectives
- ❑ Research design
- ❑ Duration
- ❑ Total Budget
- ❑ *Keep to about 250-300 words*

# Introduction

---

- Can be divided into 3 sections or the concepts of the 3 sections can be merged together
- Sections:
  - Background
  - Statement of the problem
  - Literature review
  - Justification of the study

# Problem Statement

---

Paragraph 1: Nature, magnitude and distribution of the problem

Paragraph 2: What is known (attempts done before)

**Paragraph 3: What is unknown**

Paragraph 4:  
Expectation

Tadesse A (PhD)

- Objectives
- Design
- Population
- Site
- Sampling
- Tools
- .....

# Statement of the problem

---

## Includes:

- Brief description of the socio-economic & cultural characteristics
- Overview of health status and health care system
- The **nature** of the problem; and the discrepancy

## Importance:

1. Is the **foundation** for the further development of the research proposal (research objectives, methodology, work plan, budget, etc.).
2. To **find information and reports** of similar studies.
3. Point out **why the proposed research on the problem should be undertaken**
  - Magnitude and severity could be justification

# Statement of the problem

---

- Information that should be included in the statement of the problem
  - 1. **Socio-economic and cultural** characteristics
  - 2. **Nature of the problem :**
    - **Discrepancy, size, distribution and severity** of the problem.
  - 3. **Major factors that may influence the problem**
  - 4. **Any solutions attempted to solve the problem**
  - 5. **Type of information** expected
  - 6. A short list of **definitions** of crucial concepts

# **Statement of the problem**

---

- Concise description of the nature of the problem- what is it, magnitude, distribution (who, where, when), severity and consequences.
- Systematically elucidate why the proposed research should be undertaken.
  - Brief description of any attempts to solve the problem in the past- successes, failures and challenges.
  - Provide convincing argument that available knowledge is insufficient to solve the problem under study.
- Describe the significance of the proposed study- what you hope to achieve with the study results

## Statement of the problem

---

- **Move from general to specific**
- **Engage your reader:** answer the questions, "What did you do?" "Why should I care?"
- **Make clear the links** between problem and solution, question asked and research design, prior research and yours.
- **Be selective**, not exhaustive, in choosing studies to cite and amount of detail to include. (In general, the more relevant an article is to your study, the more space it deserves and the later in the introduction it appears.)

# Literature review

---

- A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic
  
- Literature reviews are secondary sources, and do not report new or original work

# Types of literature review

---

- Narrative literature Review
- Systematic Literature Review
- Meta-analysis
- Meta-synthesis

# Literature review

---

- Search widely
  - Electronic search engines (Pubmed, HINARI,...)
  - Grey literature
- Reference all literature that you refer in your review
- Evaluate the relevance of the literature to your study
- Include information directly relevant to your study
- There should be logical sequence in writing literature review
- Be concise

# Literature Review

---

- ❑ Entertain confounding factors and other related matter only pertaining to the current research

**Avoid repetitions and lengthy statements.  
(Max 4-5 pages)**

# Purposes of Literature Review

---

- It **prevents duplicating work**
- It shares with the reader the results of other studies that are closely related to the study being reported.
- This may assist you in refining your statement of the problem.
- It relates a study to the larger, ongoing discussion in the study about a topic, filling in gaps of the study
- It provides a framework for establishing the importance of the study, as well as a benchmark for comparing the results of a study with other findings

# Purposes of Literature Review

---

- It "frames" the problem earlier identified.
- **Become more familiar with the various research approaches** that might be used in your study
- **Convincing arguments**

## Purpose of literature review



# Advantage

---

- Avoid reinventing the Wheel/avoidance of duplicating previous work
- Learn the gaps
- Learn the various methods used
- **An opportunity to develop professional confidence in the field of study!!**

# Steps of literature review

---

- **Step 1:** Begin by identifying key words or phrases useful in locating materials in an academic library at a college/university and/or websites
- **Step 2:** With these key words or phrases in mind, next connect to the internate and start searching
- **Step 3:** You would initially try to locate about 2000 reports of research in articles or books related to research on your topic.
- **Step 4:** Using this initial group of articles, you would then look at the articles and select those central to your topic. In the selection process, you would look over the abstract and skim the article or chapter.

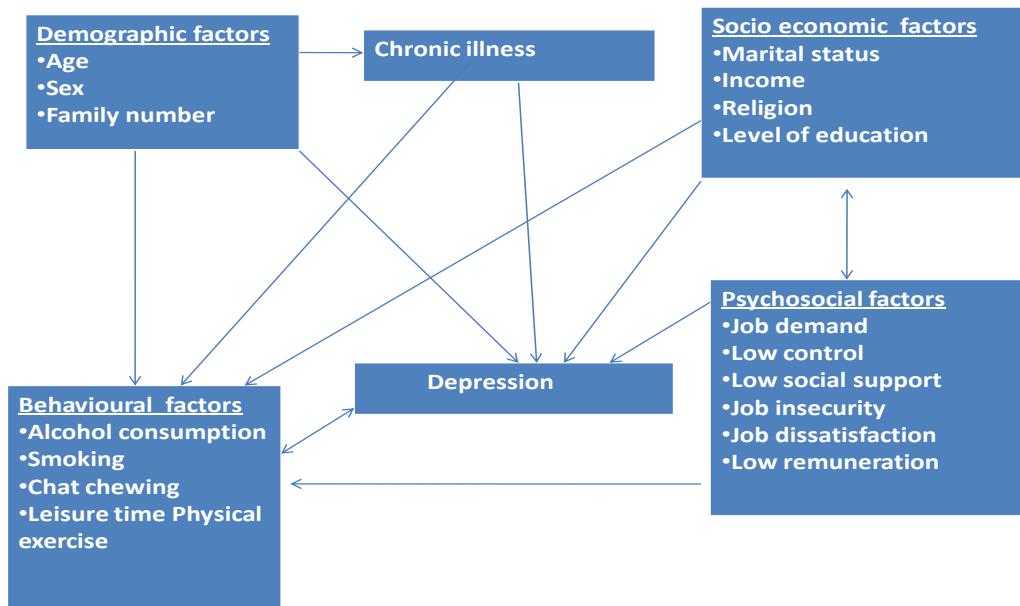
# Steps of literature review

---

- **Step 5:** As you identify useful literature, you may begin designing the literature map on your topic
- 
- **Step 6:** Organize the literature into your literature map.
  - Continue to draft summaries of the most relevant articles.
  - Summaries are then combined into the final literature review that you write for proposal.

# Conceptual framework

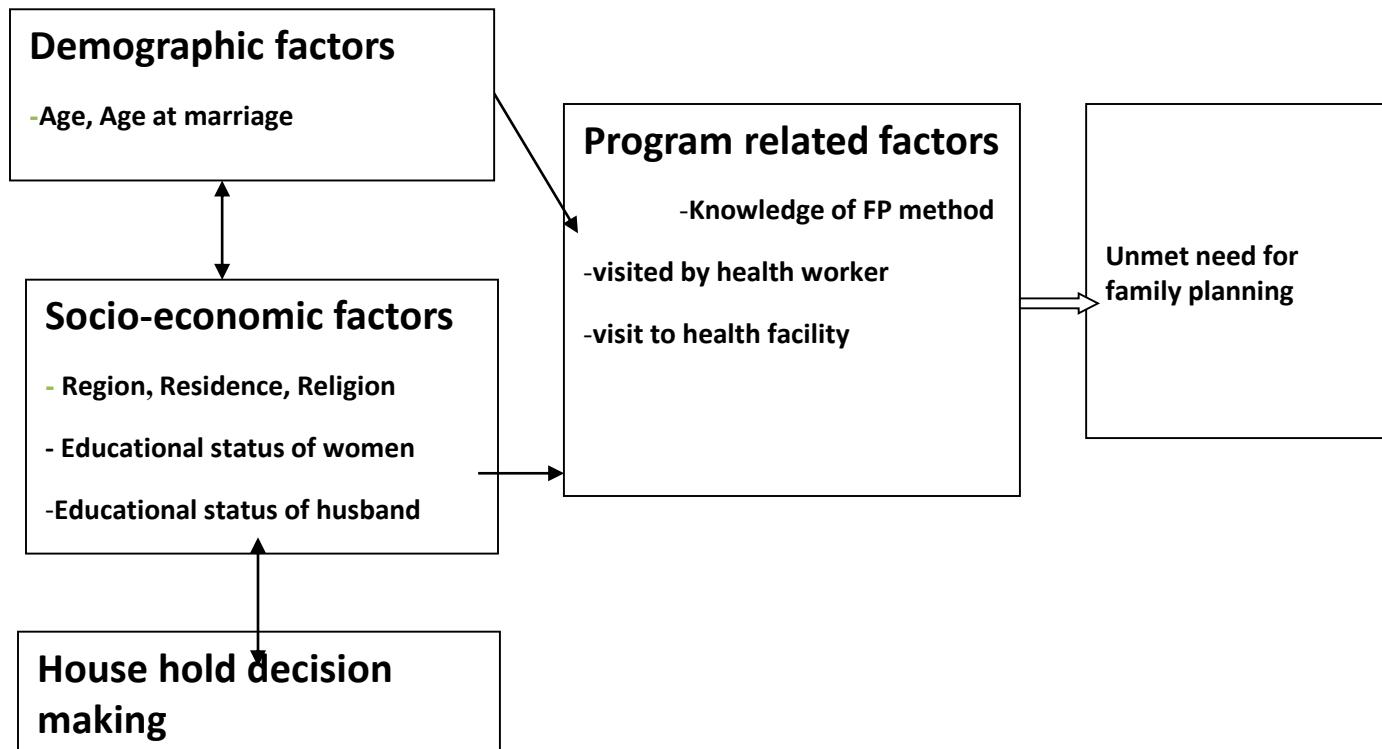
- Schematic presentation which shows the inter-relationship between the dependent variable and possible explanatory variables
- The proximal and distal variable would be presented in a meaningful way



Comment on it!

# Conceptual framework

---



- ❑ Distal factors
- ❑ Proximal factors
- ❑ Immediate factors

# Justification of the Study

---

- Questions addressed before writing the objectives of the study:
  - Are there gaps in evidence?
  - Will results influence programs, methods, and/or interventions?
  - Will results contribute to the solution of the problems?
  - Will results influence the decision making of organizations or companies?
  - What will be improved or changed as a result of the research?
  - How will results of the study be implemented, and what innovations will come about?
- In some cases, significant of the study can be included

# Assignment 2

---

- For the proposed topic of research develop;
  - Statement of the problem
  - Literature review
  - Justify the proposed research



# **Introduction to EndNote**

Tadesse A (PhD)



**Aug 2018**

# Outline

---

- Objectives
- Introduction
- Referencing
- Referencing systems
- Citing references
- Bibliographic software for reference and citation  
(EndNote)

# Objectives

---

- The session should help the trainee to:
  - describe the importance of referencing
  - Describe the different style of reference
  - Use the different searching mechanisms
  - Able to cite the sources appropriately
  - Develop library for the materials used

# Publications (Nutrition related)

1. Low Birth Weight & Associated Factors Among Newborns in Gondar Town, North West Ethiopia: Institutional Based Cross-Sectional Study, 2014; 4(2): 74-80
2. Prevalence of hypertension and its associates among Adults in Gondar, North-West Ethiopia: a Community Based Cross-sectional Study, *BMC Cardiovascular Disorders* 2012, **12**:113
3. Timely initiation of complementary feeding practice and associated factors among mothers of children aged from 6 to 24 months in Axum town, north Ethiopia: **International Journal of Nutrition and Food Sciences**: 2014; 3(5): 438-442
4. Low knowledge and practice on prevention of hypoglycemia among diabetic patients in South Gondar, Northwest Ethiopia: Institution based cross-sectional study
5. Prevalence of zinc deficiency and its association with dietary, serum albumin and intestinal parasitic infection among pregnant women attending antenatal care at the University of Gondar Hospital, Gondar, Northwest Ethiopia: *BMC Nutrition* 2015, 1:31
6. Malnutrition in Healthy Individuals Results in Increased Mixed Cytokine Profiles, Altered Neutrophil Subsets and Function
7. Undernutrition and associated factors among adults living with Human Immune Deficiency Virus in Dembia District, northwest Ethiopia: an institution based cross-sectional study

# Publications (Mental Health)

---

1. Mental distress among university students in Ethiopia: a cross sectional survey. *The Pan African Medical Journal*. 2013;15:95
2. Depression among women with obstetric fistula, and pelvic organ prolapse in northwest Ethiopia: *BMC Psychiatry* 2013, **13**:236
3. Antipsychotic Medication Induced Movement Disorders: The Case of Amanuel Specialized Mental Hospital, Addis Ababa, Ethiopia. *American Journal of Psychiatry and Neuroscience*. Vol. 2, No. 5, 2014, pp. 76-82.
4. Community Perception towards Mental Illness among Residents of Gimbi Town, Western Ethiopia
5. Prevalence and Associated Factors of Antenatal Depression among Women Attending Antenatal Care Service at Gondar University Hospital, Northwest Ethiopia
6. Prevalence and correlates of depression and anxiety among patients with HIV on follow up at Alert Hospital, Addis Ababa, Ethiopia, *BMC Psychiatry* (2016) 16:368

# Introduction (1)

---

- Citation is the means of referring the source from which the information has been taken
- There are different means of citation
  - Manually
  - Using software
    - EndNote
    - Zotero
    - Reference manager

# Why is it so important?

---

- Established academic practice
- Shows where you got your information
- Allows other researchers to trace your sources of information quickly and easily
- Acknowledges the work of other researchers
- Protects you against accusations of plagiarism

# What to refer to

---

- References are useful to illustrate facts
- When you use references, it will support your description as having testimony
- It is usually used in
  - the introduction session as literature review,
  - The methods session on how you use past instruments including study area and sampling
  - The discussion part as testimony

# What to take from an article

---

- Once if an article is read, try to criticize it first (strength and limitations of the article).
- Criticize how the objective is assessed (methods session) and the outcome is discussed
- If you believe it is good, take the idea the paper has provided to science
- Write the idea based on your own words from your own perspective

# What should we cite

---

- Word by word copying is plagiarism, so try not to use such a way
- Any written description in the introduction of any reference is not reference of that article
- Reference is a fact of that paper, only if it is demonstrated in the result and discussion session of the reference.
- Most of the reference of any article should be taken from a peer reviewed article (non-abstract, published papers)
- 2/3 of the reference should be from peer reviewed article

# Selecting a reference

---

- First read title of a journal, at least the dependent or independent variable should be found
- If word within title present, continue to abstract session
- If abstract is related to your research area, try to find the whole part of the paper and read for your understanding
- Use **HINARI**
  - Username:
  - Password:

# Which system to use?

---

- Various different systems (or ways) in use
- For a thesis/dissertation, your institution probably may advise you on which system you should follow!  
Example: UoG → **Vancouver method**
- If you are writing a paper for an academic journal, use the house style that they specify
- In general the following two general recommended schemes are used:  
**1. Harvard,**                           **2. Numeric**
- Harvard probably most suitable for SS subjects

# What is EndNote?

---

- It is a computer software that could arrange us a reference.
- A reference used during manuscript writing.
- References could be written the way we want to write.
- We right references specifically, and endnote will use reference writing style specific to what we selected

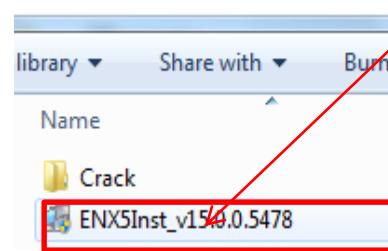
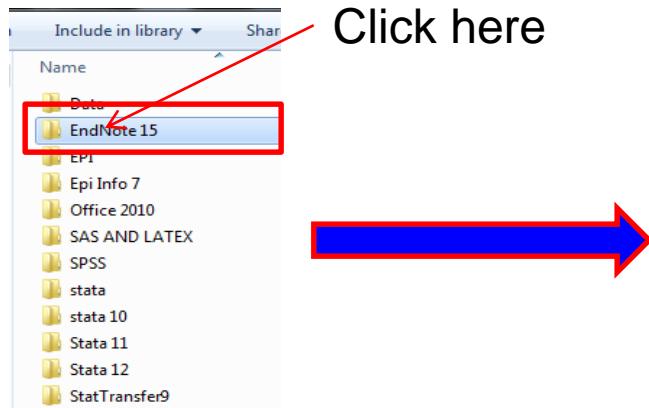
# Use of *EndNote*

---

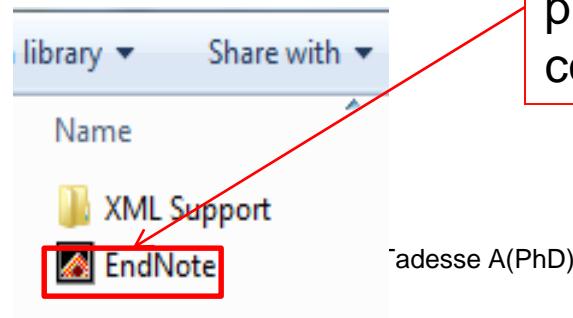
- **EndNote** is an **online search tool**
  - it provides a simple way to search online bibliographic databases and retrieves the references directly into EndNote.
- **EndNote** is a **reference database (library)**
  - it specializes in storing, managing, and searching for bibliographic references in your private reference library.
- **EndNote** is a bibliography and **manuscript maker**
  - it formats citations in **Microsoft® Word** with the (Cite While You Write™) feature.

# Installation

- To install, open the folder containing the software



- After installation, go back to the folder containing EndNote and open crack

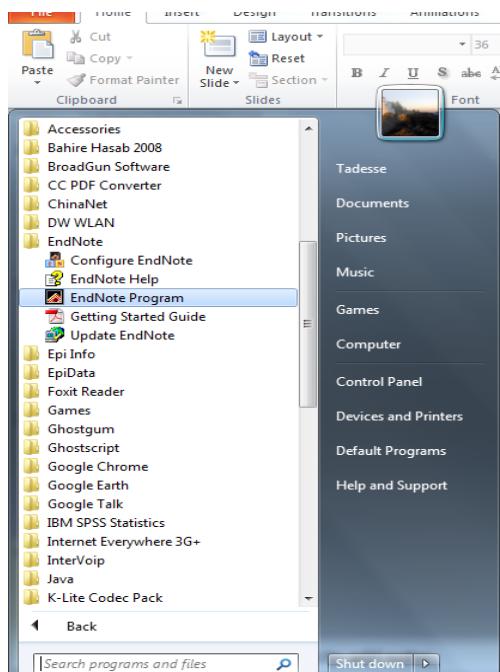


Copy EndNote and past it to  
program file in C/ within  
containing the folder EndNote

# Starting Endnote

1. From the *Start* menu, choose *Programs*, select *Endnote*, and then choose the *Endnote Program*.

*Start* → *Programs*, → *Endnote* → *Endnote Program*



Tadesse A(PhD)

# Starting...

---

2. Select *Open an existing EndNote library, if there is existing file.*
3. If it is for the first time, select new and write name of the library

What is a library?

- A library is list of journals, books, website citations ready to be cited in a manuscript

# This is an example of a library

The screenshot shows the EndNote software interface. On the left is a sidebar with categories like 'My Library', 'My Groups', 'Online Search', 'EndNote Web', and 'Find Full Text'. The main area displays a list of references with columns for Author, Year, Title, Journal, Ref Type, URL, and Last Updated. A search bar at the top includes fields for 'Quick Search' and 'Search Remote Library' with options for 'Match Case' and 'Match Words'. The bottom of the screen shows a search interface with dropdowns for 'Year', 'Title', and 'And' operators, and a text input field containing 'HIV/AIDS in Africa'.

	Author	Year	Title	Journal	Ref Type	URL	Last Updated
All References	Sorsdahl, K...	2010	Perspectives towards mental illness i...	AIDS Care	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
Unfiled	McFarland, ...	2010	HIV/AIDS in the Middle East and Nor...	AIDS	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
Trash	Lasry, A.; Ca...	2011	Allocating funds for HIV/AIDS: a desc...	Health Poli...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Robberstad, ...	2010	The health related quality of life of pe...	Cost Eff Re...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Rohleder, P.;...	2010	HIV/AIDS and disability organisation...	AIDS Care	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Garcia-Jard...	2010	Postmortem findings in HIV/AIDS pat...	Trop Doct	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Mamotte, N.;...	2010	Convergent ethical issues in HIV/AID...	BMC Med ...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Shisana, O.; ...	2010	Gender and poverty in South Africa in...	J Womens ...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Letwaba, R.;...	2010	Maintaining effective psychosocial tr...	J Dev Behav...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Zuniga, J. M.	2010	HIV/AIDS workforce development in ...	J Int Assoc ...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Grange, J.; A...	2010	Tuberculosis in association with HIV/...	Int J Gynae...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Seeley, J.; D...	2010	The effects of HIV/AIDS on rural com...	Trop Med In...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Long, L.; Fox...	2010	The high cost of second-line antiretro...	AIDS	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Wendland, C...	2010	Health Electives in Africa and the Dut...	Virtual Ment...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Cloete, A.; S...	2010	Challenges Faced by People Living ...	AIDS Res T...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Gedzedha, ...	2010	Should routine serological screening ...	S Afr Med J	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	Gilbert, L.; W...	2010	'My biggest fear was that people wou...	Health Soc ...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014
	King, R. C.; ...	2010	Remodeling pharmaceutical care in ...	Int J Health ...	Journal Arti...	http://www.ncbi.nlm...	6/28/2014

# Adding a reference to a library

---

- There are various ways to add references to an EndNote library:
  1. Type the reference information into the Reference window **manually**.
  2. Connect to an **online bibliographic** database and retrieve references directly into EndNote.
  3. Import text files of references that have been downloaded from online bibliographic databases or **CD-ROMs**

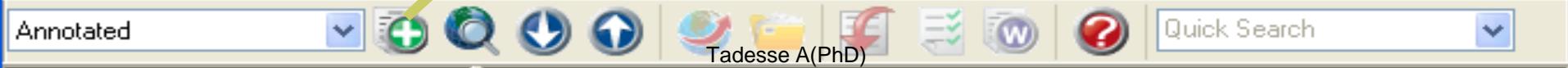
# 1. Manual reference writing.

---

- We type reference manually when the article is not found on the online (non-published or not cited on Medline/ PubMed)

How to do it:

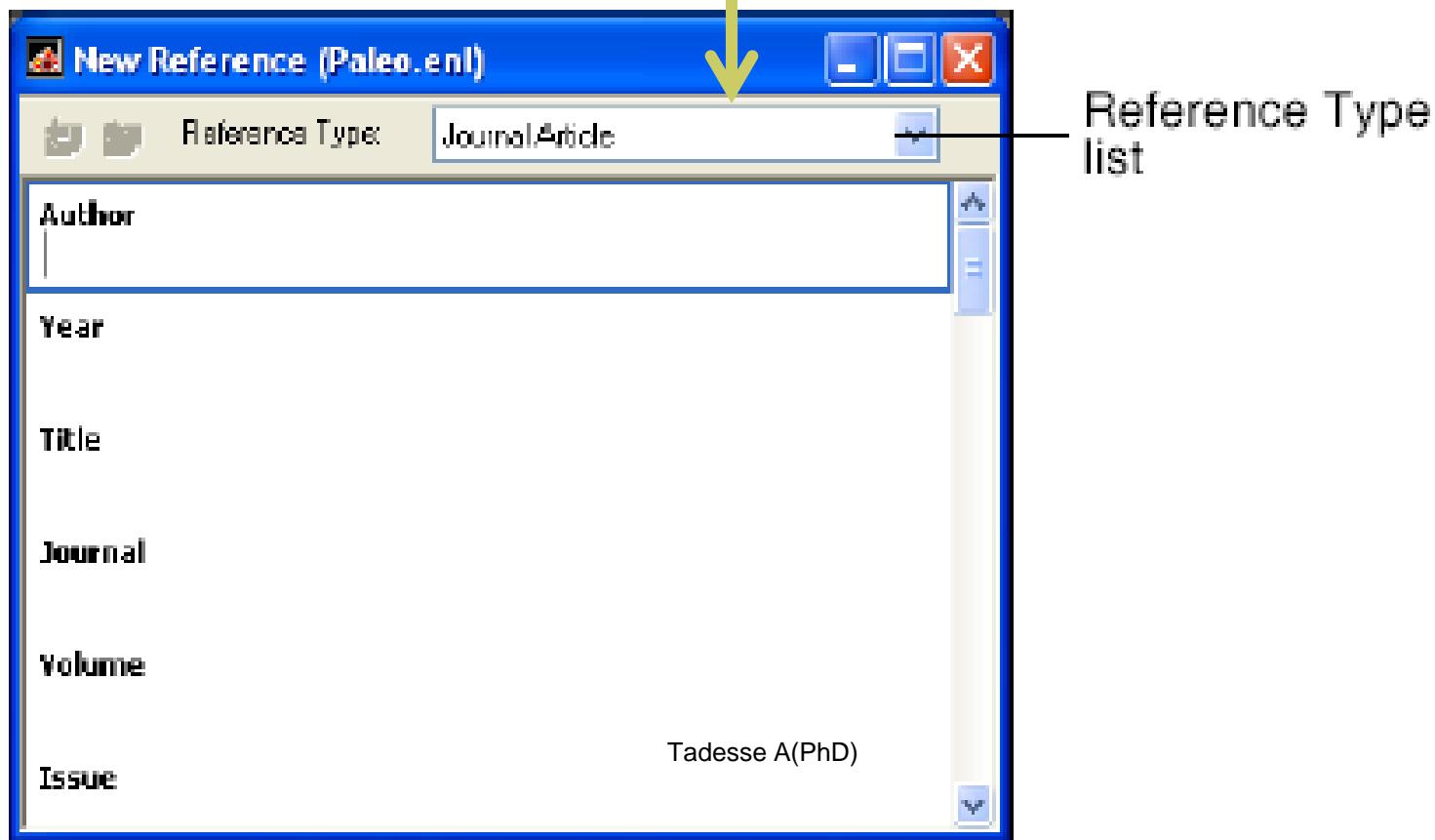
- Once a library is open, you can add a new reference :
  1. From the *References menu*, choose *New Reference* (*CTRL+N*).
  2. You can also click this figure from the main toolbar



# This menu comes

---

The reference may be a book, journal, internet browse, etc  
select by drop downing the reference type



This menu comes...

---

## Author's names.

### 1. Author names can be entered

- Last or middle name of each author is written first
- Individual author names *must* be written on **one author per line**.
- Make a **comma** “ , ” after the written name
- Then write the first name or **first alphabetical letter** of first name

#### □ Example

### writing

- Eg      Kebede, Y  
              Getachew, A

Written in library

**Real name**

Yigzaw Kebede  
Asminew Getachew

# This menu comes...

---

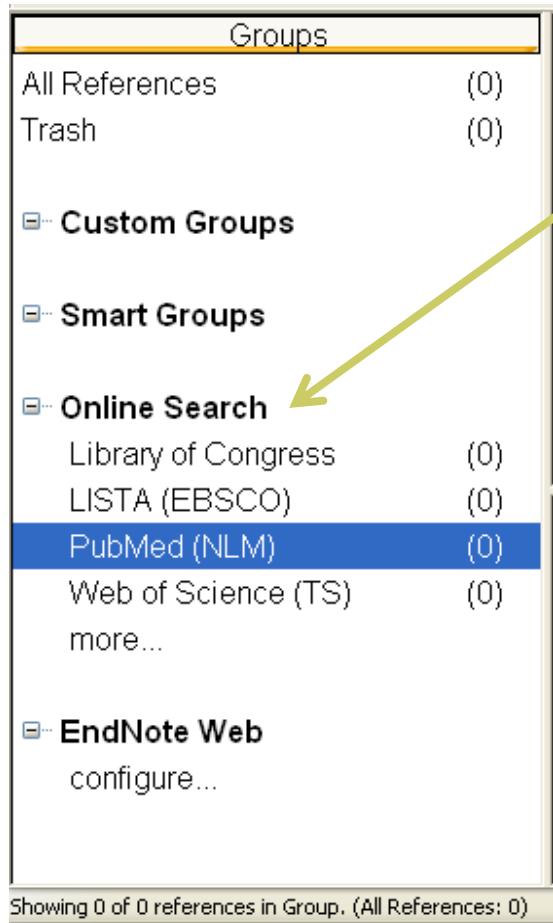
- The next step is to write
  2. the year,
  3. the title,
  4. volume,
  5. issue,
  6. page number,
- but don't use comma, parenthesis or other  
(other than what you write on the name of authors)  
(endnote will complete itself)
- After completing, close by clicking the **close** button



## 2. Online Retrieval of reference

---

It has **five steps**



1. From the *groups* chose  
*online search*

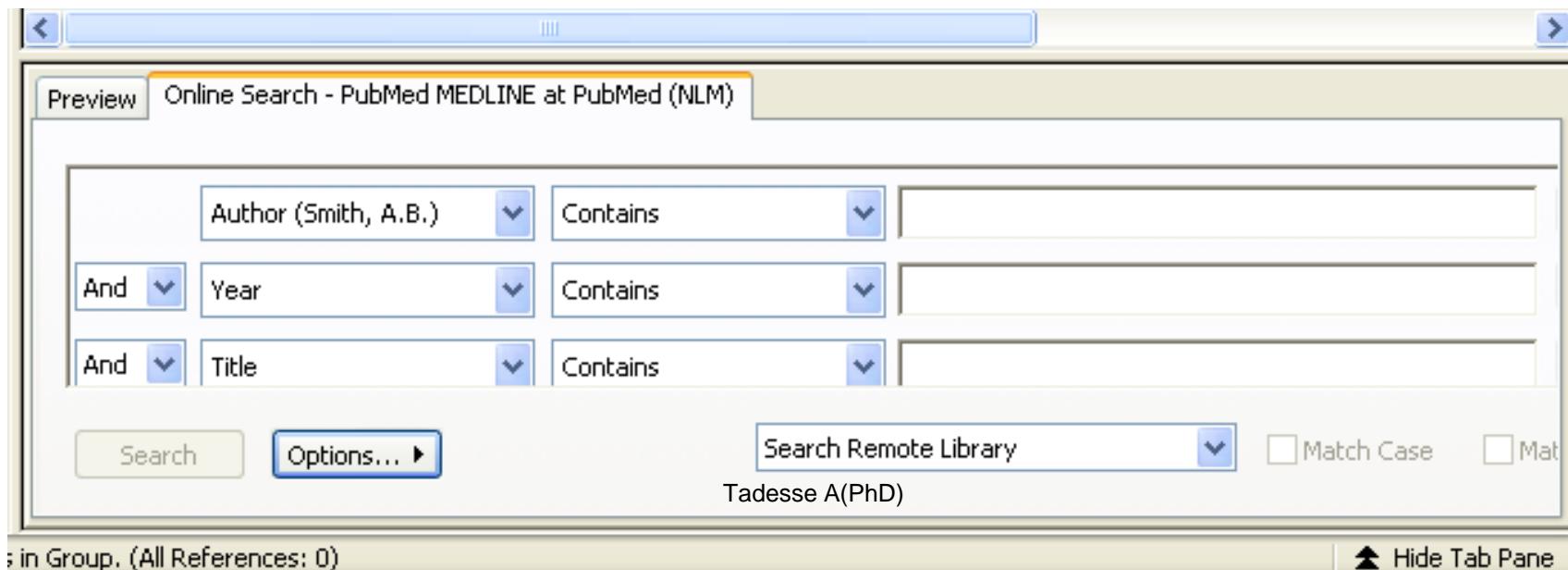
2. Select the database that  
you would like to search, eg  
*pubmed (NLM)* and click  
*Connect.*

Tadesse A(PhD)

# Online Retrieval...

---

- When the connection has been established, an empty Retrieved References window opens
- EndNote's Search window appears, ready for you to enter a search.



# Online Retrieval...

---

3. Enter your search request into EndNote's Search window, and click *Search*.

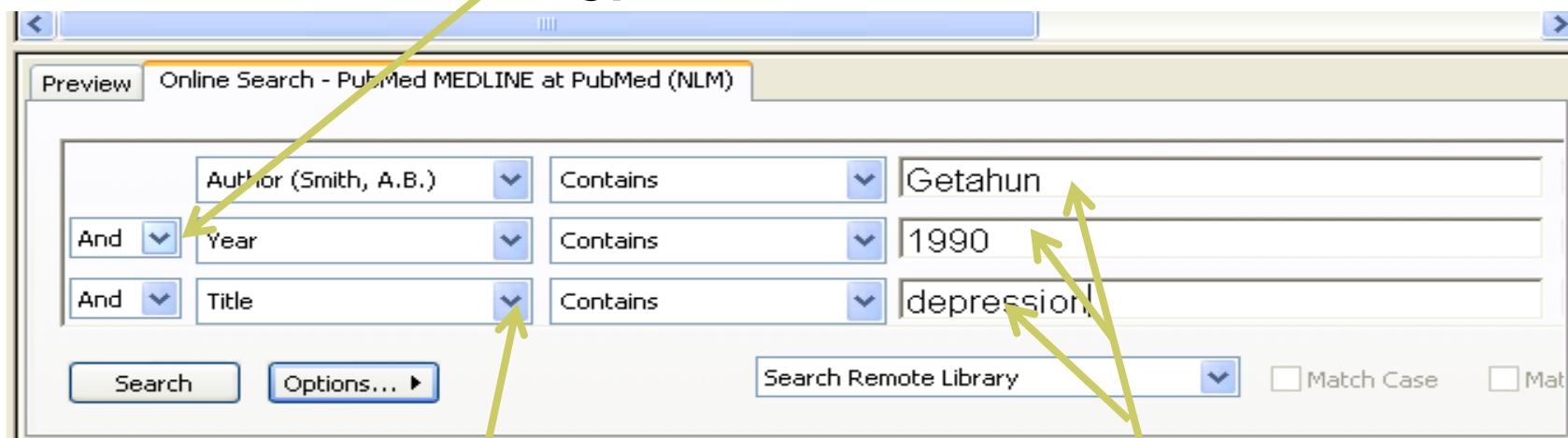
After establishing a connection:

- A. Enter your search term(s) into EndNote's Search window.
- B. Choose the appropriate search options.
- C. Click *Search to send the search request to the remote database.*

# Online Retrieval...

- After entering your search key words it may seem as figure below

and  
or



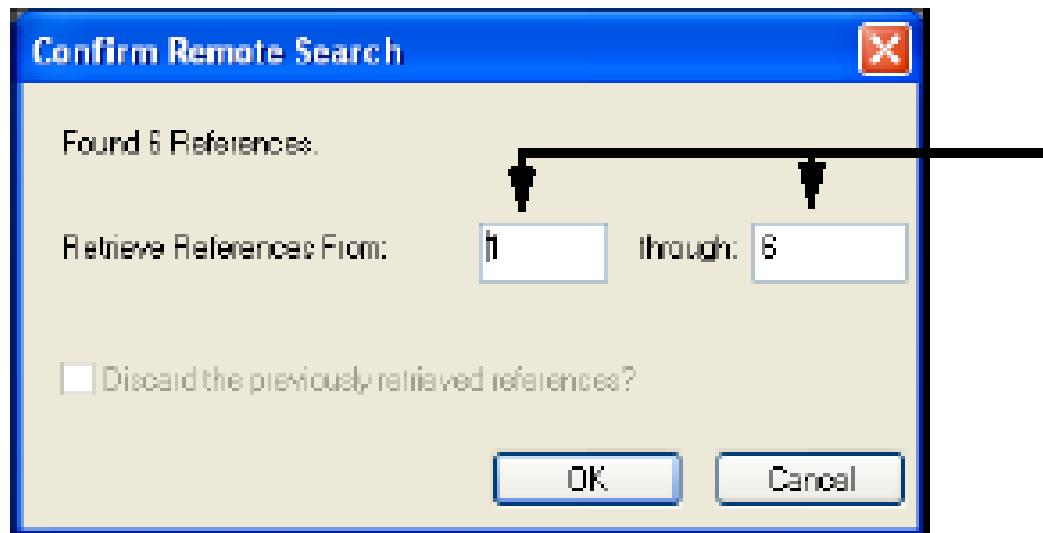
Author  
Title  
Year  
Any field

Tadesse A(PhD)  
Clicking here you are able to choose  
Place of the searched item

# Online Retrieval....

---

4. EndNote searches the database and displays the number of references that were found to match your search criteria. Click *OK* to retrieve the references.



Change one or both of these numbers to specify the range of references to be downloaded.

5. Once the references appear in the Retrieved References window, you may browse through them and these references you assume are not useful should be transferred to the 'trash'

The screenshot shows the EndNote X2 interface with a blue header bar containing the title 'EndNote X2 - [Sample\_Library]' and standard menu options: File, Edit, References, Groups, Tools, Window, Help. Below the header is a toolbar with icons for New, Open, Save, Print, and Search, along with a 'Quick Search' field. The main window is divided into two sections. On the left is a sidebar titled 'Groups' with a tree view. It shows 'All References' (219), 'Trash' (0) which has a red arrow pointing to it, 'Custom Groups' (Excavation, Fossils, Literature), 'Smart Groups' (Eocene), and 'Online Search' (Library of Congress, LISTA (EBSCO), PubMed (NLM) which is highlighted with a red box, and Web of Science (T...)). On the right is a table view of references. The columns are labeled: Groups, Author, Year, Title, Journal, and Ref Type. The table lists several entries, such as Tesfaye (2009), Deyessa (2009), Terasaki (2009), Deyessa (2008), Hanlon (2008), Summerfield (2007), Mogga (2006), Frost (2002), and Conroy (1978). The 'PubMed (NLM)' group in the sidebar is also highlighted with a red box.

By double clicking the reference you are able to browse details of the reference

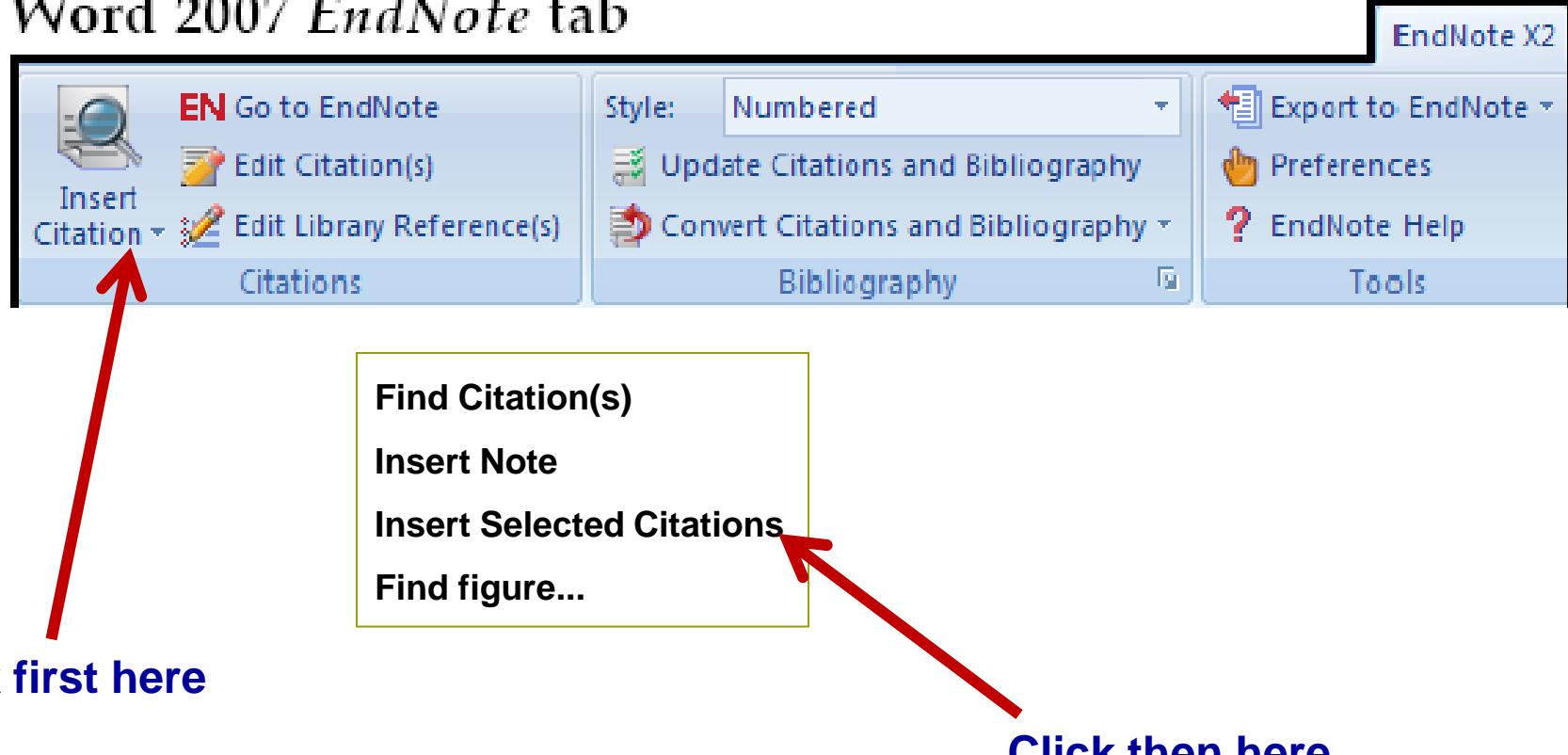
# 1. Cite While Writing

---

1. When you are ready to cite a source, position the cursor in the text where you would like to put the citation.
2. Go to your endnote library, and select the reference(s) you want to cite.
3. Go back to word office and click the ***EndNote X2*** menu in Word, and find the ***Insert Citation*** and click pull-down to find ***Insert Selected Citation(s)***.
4. Click the ***Insert Selected Citation(s)***, and the selected reference(s) will be cited in the cursor and the bibliography will be written on the reference site

# Cite...

Word 2007 *EndNote* tab



Click first here

Click then here

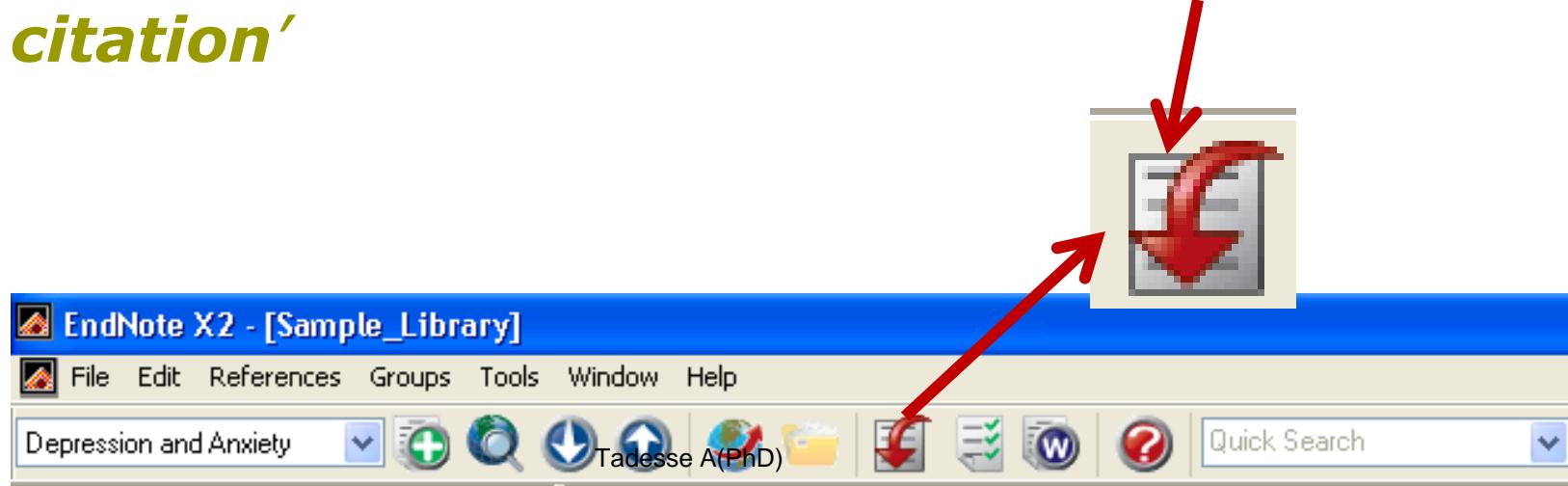
You will see the reference cited, and bibliography written

Tadesse A(PhD)

## 2. Cite While You are on EndNote

---

1. When you are ready to cite a source, position the cursor in the text (word offce) where you would like to put the citation.
2. Then shade the reference you want to cite
3. In the EndNote menu, click the '***Insert citation***'



## 2. Open the Word document and position the cursor

---

- .....reported that for every maternal death in developing countries, at least 15 other women are incapacitated or handicapped, and about 15 million women develop long term disabilities such as obstetric fistula, sever anaemia, pelvic inflammatory diseases, reproductive tract infection and infertility .

### 3. Highlight the desired reference(s) in your EndNote library

disabled

0	Author	Year	Title	UR
	Abdulahi	2001	Burden of disease analysis in rural Ethiopia	http
	Abu Habib	1995	"Women and disability don't mix]" Double discrimination and disabled women's rights	http
	Alem A. Kebede...	2004	Comparison of Computer assisted Scan and Clinical Diagnosis of Major Mental Di...	
	Beyero	2004	Mental disorders among the Borana semi-nomadic community in Southern Ethiopia	http
	Cardol	2002	Beyond disability: perceived participation in people with a chronic disabling condition	http
	Carvill	2002	People with intellectual disability, sensory impairments and behaviour disorder: a ca...	http
	CSA	1994	Population and Housing Census of the Federal Government of Ethiopia	
	Deyessa N. B...	2007	Determinants for depression among women in rural Ethiopia: A community based st...	
	Deyessa N. Be...		Violence against women and depression in rural Ethiopia	
	Disler	1986	The prevalence of locomotor disability and handicap in the Cape Peninsula. Part III. ...	http
	Disler	1986	The prevalence of locomotor disability and handicap in the Cape Peninsula. Part II. ...	http
	Disler	1986	The prevalence of locomotor disability and handicap in the Cape Peninsula. Part I. T...	http
	Ellsberg	1999	wife Abuse Among women of child bearing Age in Nicaragua	
	Elman	1997	Disability pornography: the fetishization of women's vulnerabilities	http
	Eshete	Jan...	"Women in Ethiopia problem and prospects." Paper presented at seminar on Gend...	

Alem A. Kebede D. Shibre T. Negash A. Deyessa N. 2004. Comparison of Computer assisted Scan and Clinical Diagnosis of Major Mental Disorders in Butajira, Rural Ethiopia. *Ethiop Med J* 42:137-143.

Showing 40 out of 40 references. Tadesse A(PhD) Hide Preview

# Final result

---

- .....reported that for every maternal death in developing countries, at least 15 other women are incapacitated or handicapped, and about 15 million women develop long term disabilities such as obstetric fistula, sever anaemia, pelvic inflammatory diseases, reproductive tract infection and infertility (Alem A. Kebede D. Shibre T. Negash A. Deyessa N. 2004).

## Reference

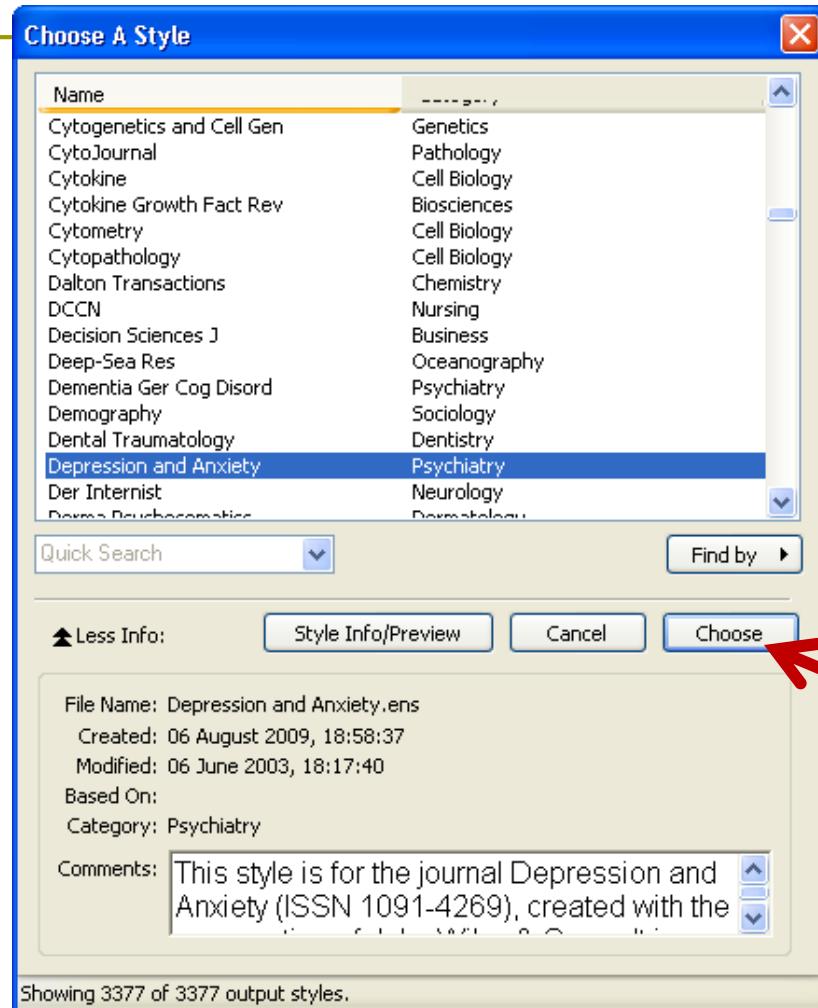
Alem A. Kebede D. Shibre T. Negash A. Deyessa N. 2004. Comparison of Computer assisted Scan and Clinical Diagnosis of Major Mental Disorders in Butajira, Rural Ethiopia. *Ethiop Med J* 42:137-143.

# Changing output style of reference

---

1. Open the EndNote libraries
2. Select and click “**format bibliography**”
3. Select another style....
4. Thousands of options of output styles will come
5. Select the output style you wish

# Select type of style



Choose the style you wish

# Moving part of a paragraph

---

- The good thing about endnote is when you move a paragraph having a citation of reference, the bibliography is moved to
  
- You are able also to change from any format to other according the journal you want to send

# The End



# Session II: Formulation of objectives



Tadesse Awoke (PhD)

- General objective
- Specific objectives

*“The formulation of a problem is often more essential than its solution”*

- Albert

*Einstein*



# Learning Objectives

---

At the end of this session the student will be able to:

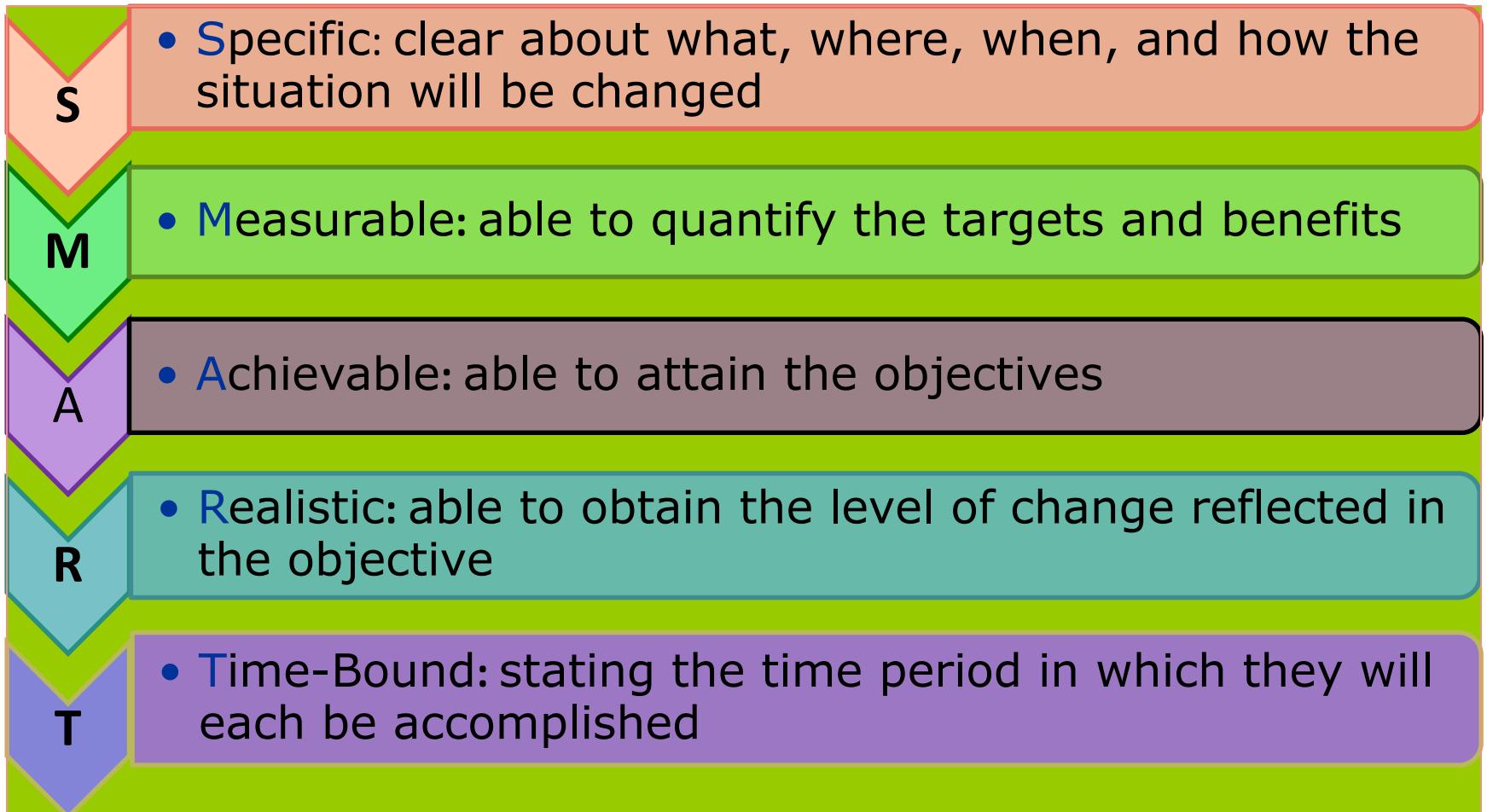
- Develop research objectives of different forms
  
- Formulate research hypothesis

# Research Objectives

---

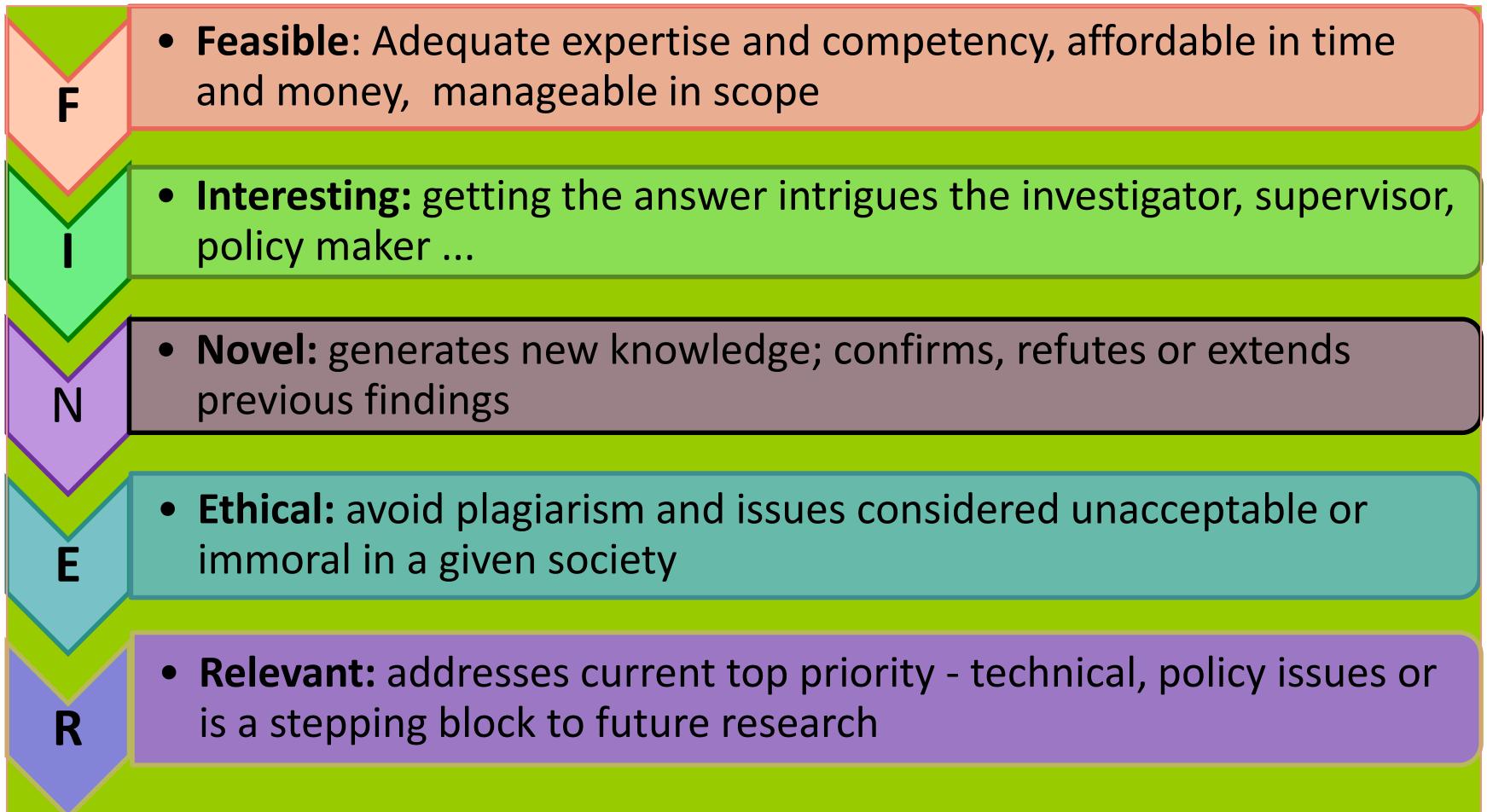
- Once you understand what is known and what is not known, determine what can be answered through primary research
  - Avoid trying to answer too many questions in one study
  - Lots of questions require multiple studies or complex designs
  -
- Prioritize questions as “primary” and “secondary”
  - Try to stick to 1-2 primary questions and never compromise the primary questions

# Criteria for a good research Objective



**SMART criteria for a good research question and objective**

# Criteria for a good research Objective



**FINER criteria for a good research question and objective**

# Objectives Formulation

---

- Commonly, research objectives are classified into **general objectives and specific objectives.**
- **General objective:**
  - summarizes what is to be achieved by the study
  - should be clearly related to the statement of the problem
- **Specific objectives:**
  - are logically connected parts of the general objective
  - **focus the study on the essentials**
  - **direct the design of the investigation**
  - **orient collection, analysis and interpretation of the data**
  - Use action verbs

# Criteria for setting research objectives

---

- **Focused**, each covering a single point
- **Ordered** in a logical sequence
- **Realistic** and feasible to answer
- **Operational**, using *action verbs* such as:
  - determine      - verify                        -identify
  - describe        - assess                        - compare
  - calculate       - establish                    -explore
- **Measurable** outcomes at the end of the research

# Importance of developing objectives

---

- Avoid the collection of data which are not strictly necessary
- Properly formulated specific objectives facilitate the development of research methodology and help to orient the collection, analysis, interpretation and utilization of data.
- Helps for evaluating the project

# Format of Research Objective

---

**Research objectives can be stated as:**

- A) **Questions:** “The objectives of this study are to answer the following questions ...”
- B) **Positive sentence:** “The objectives of this study are to determine ...”
- C) **Hypothesis:** “The objective of this study is to verify the following hypothesis...”

## What formats cont...

---

- The format chosen depends on the type of study that will be undertaken
- If the study is descriptive or exploratory in nature, then objectives are stated in the form of questions or positive sentences.
- If the researcher knows enough to make predictions concerning what s/he is studying, then hypotheses may be proposed.

# *Research questions Vs hypotheses*

---

- In order to develop the research design, the research topic often has to be changed to a research question or hypothesis
- ***Research questions:*** formulated when insight into the problem being studied is not sufficient
- **Research question should:**
  - **Be focused**, each covering a single point
  - **Be ordered** in a logical sequence
  - **Be realistic** and feasible to answer
  - **Have measurable** outcomes at the end of the research
- Eg: Does post-menopausal hormone replacement therapy predispose women to develop endometrial cancer?

# Formulating Objectives

---

**Example:** Imagine you are a member of National Cigarette Smoking Cessation Group.

- One issue that the group considers is nicotine replacement therapy for people who want to stop smoking.
- The other group says advising people to stop smoking will suffice.
- The debate continued but they couldn't agree.
- Formulate a research question for the above problem

# Formulating Objectives

---

- **Explicit hypothesis:** requires sufficient knowledge of the problem to be able to predict relationships among factors which then can be explicitly tested.

E.g. Post-menopausal women who received hormone replacement therapy are more likely to develop endometrial cancer than post-menopausal women who did not receive such therapy

# Assignment 3

---

- For your proposed research topic
  - Formulate the objective using appropriate format

# Session III: Methods



Tadesse Awoke (PhD)

Aug 2018

# Learning Objectives

---

At the end of this session the student will be able to:

- Analyze the circumstances for appropriate use of the different study designs
  
- Use appropriate study methods which properly answer the research question/ objective

# Methods

---

- **Possible sub-sections of the methods:**
  - Study design
  - Study setting
  - Population
    - Source population
    - Study population
  - Inclusion/Exclusion Criteria
  - Sample size
  - Sampling strategies
  - Variables of the study
    - Dependent/outcome/Response variable
    - Independent/explanatory variables

# Methods

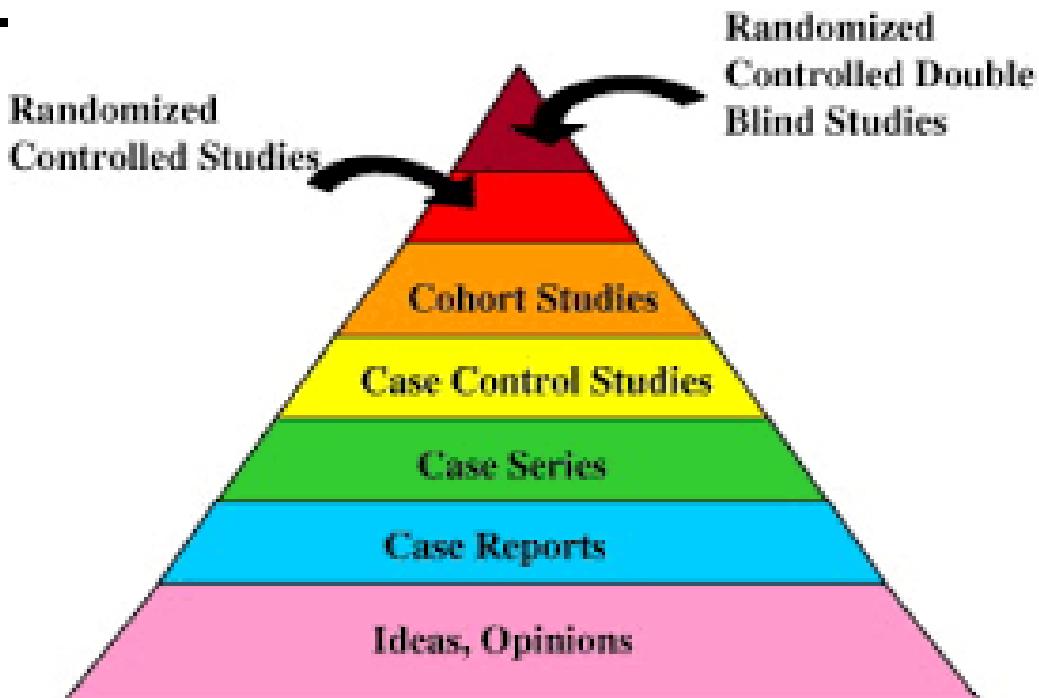
---

- Data collection:
  - what type,
  - how (tool/procedure),
  - Who responds,
  - where (geography/place),
  - when (period/time)
- Data quality control issues
- Data management:
  - coding, entering,
  - cleaning, storing,
  - recoding (software to be used)
- Data analysis: choice of statistical methods

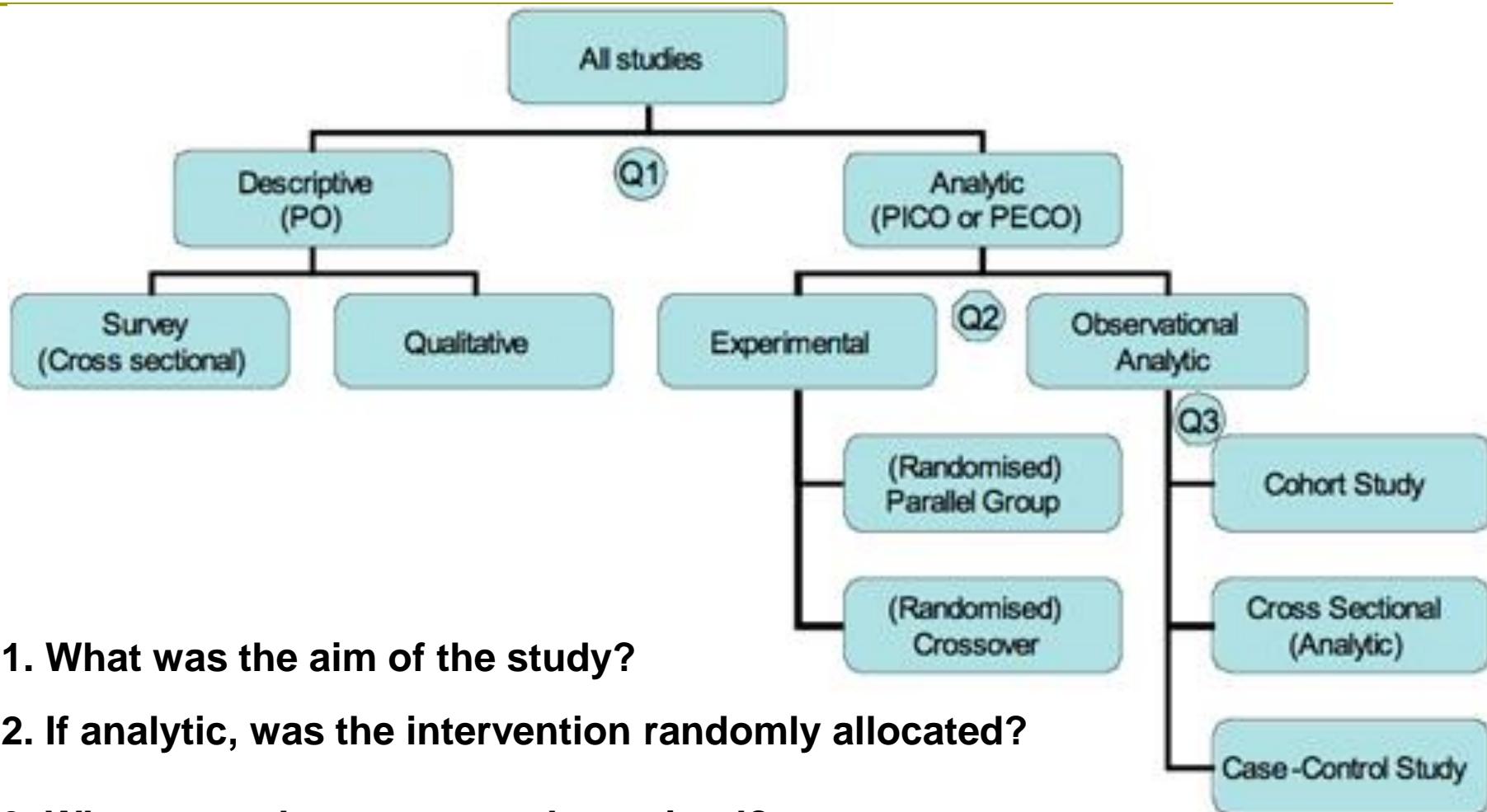
# Study design

---

- A **study design** is a specific plan or protocol for conducting the **study**, which allows the investigator to translate the conceptual hypothesis into an operational one.



# Quantitative Study Design



Tadesse A (PhD)

# Qualitative Designs

---

- ❑ Phenomenology
- ❑ Ethnography
- ❑ Case study
- ❑ Grounded theory
- ❑ Historical study

# Criteria for selecting a research design

---

- 1) The research problem
- 2) Personal experience
- 3) Audience
- 4) Resources available for the study

## Criteria for selecting cont..

---

### 1) The research problem

Research problems call for specific approaches

Eg. If the problems calls for:

- a) the identification of factors that influence an outcome
- b)) the evaluation of an intervention

Then a quantitative approach is best

- If a concept or phenomenon needs to be understood because little research has been done on it,

qualitative approach is preferred

# Criteria for selecting cont..

---

- A mixed methods design is useful when either the quantitative or qualitative approach by itself is inadequate to best understand a research problem
- The strength of both quantitative and qualitative research can provide the best understanding

## 2) personal experience

- Researcher's personal training and experience influence the choice of the approach
- An individual trained in technical, scientific writing, statistics, and computer statistical programs and familiar with quantitative journals would most likely choose the quantitative design

# Criteria for selecting cont..

---

- Individuals who enjoy writing in a literary way or conducting personal interviews, or making up-close observations may prefer qualitative approach
- The mixed methods researcher is an individual familiar with both quantitative and qualitative research

## 3) Audience

- Researchers write for audiences that will accept their research
- These audiences may be journal editors, journal readers, conference attendees, or colleagues in the field

# Quantitative designs

---

- Since quantitative studies are traditional mode of research, carefully worked out procedures and rules exist for them
- Researchers may be more comfortable with the highly systematic procedures of quantitative research

# Case report & case series

---

- A case report is a descriptive study of a single individual (case report)
- case series is a descriptive study of small group
- Ecologic Studies: The unit of analysis is the group, not the individual.

## **Example:**

- Childhood lead poisoning in communities in Massachusetts.
- Mean systolic blood pressure levels and stroke mortality rates in the Seven Countries Study.

## Cross-sectional study

---

- A type of prevalence study.
- Exposure and disease measures obtained at the individual level.
- Single period of observation.
- Exposure and disease histories collected simultaneously.
- Both probability and non-probability sampling used.
- **Example:** Prevalence of congenital malformations across maternal age groups.

# Uses of cross-sectional study

---

- ✓ Hypothesis generation
- ✓ Intervention planning
- ✓ Estimation of the magnitude and distribution of a health problem

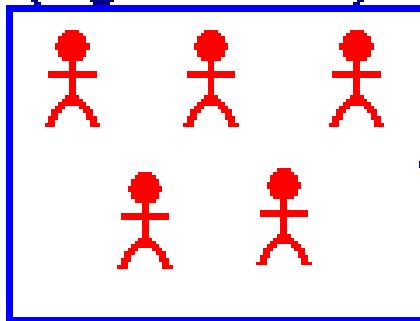
# Cohort Studies

---

- In a cohort study, subjects with an exposure to a causal factor are identified and the incidence of a disease over time is compared with that of controls (persons who do not have the exposure).
  
- In a longitudinal study, subjects are followed over time with continuous or repeated monitoring of risk factors or health outcomes, or both.

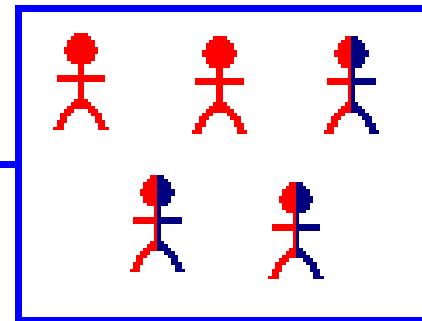
# Cohort Studies

**Group of interest  
(e.g. smokers)**

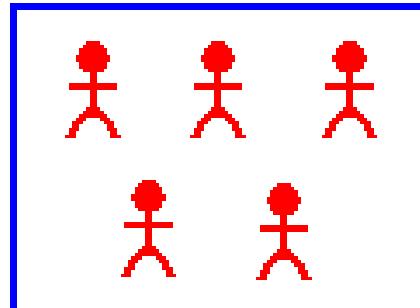


**Follow**

**over time**

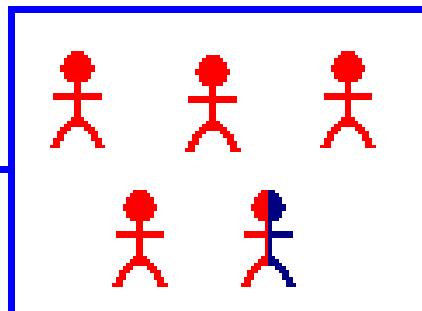


**Comparison group  
(e.g. non-smokers)**



**Follow**

**over time**



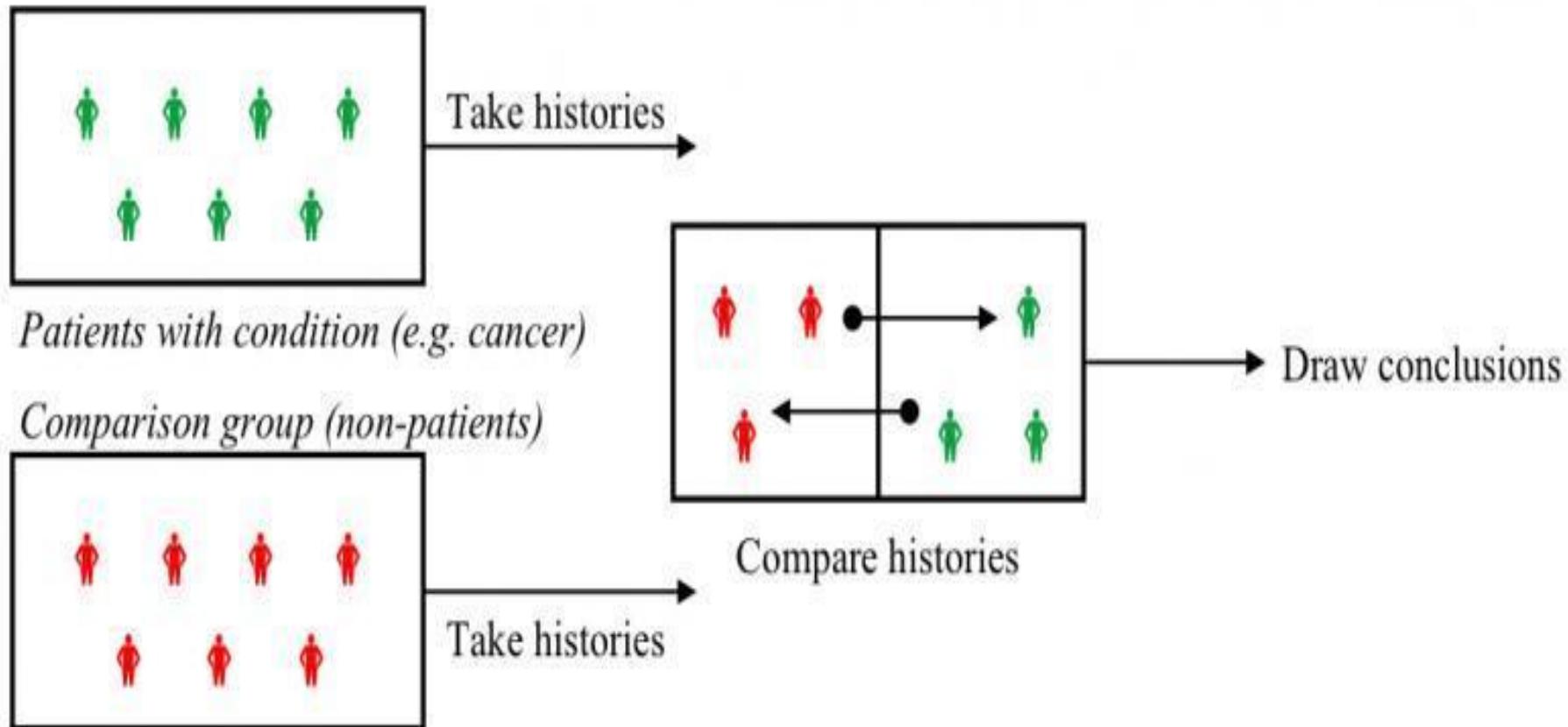
**Compare  
outcomes**

# Case-control studies

---

- Case-Control Studies identify existing disease/s and look back in previous years to identify previous exposures to causal factors.
  - **Cases are those who have a disease**
  - **Controls are those without a disease**
- Analyses examine if exposure levels are different between the groups.

# Case-Control studies



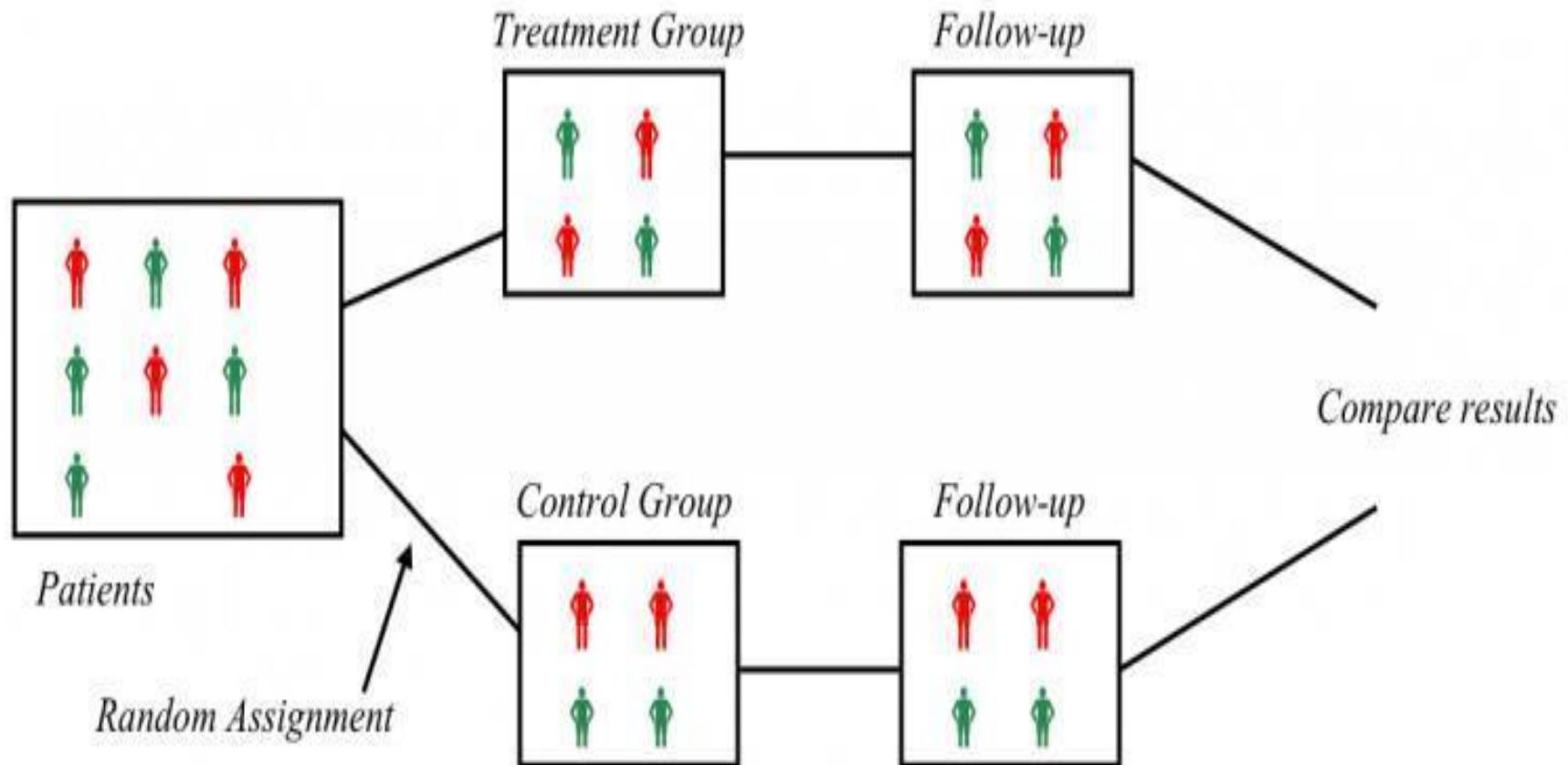
## Randomized Controlled Studies

---

- There are two groups, one treatment group and one control group
- The treatment group receives the treatment under investigation, and the control group receives either no treatment (placebo) or standard treatment
- Patients are randomly assigned to all groups
- Randomized controlled trials are considered the “gold standard” in medical research

# Randomized Controlled Studies

---



# Other designs

---

**Time series** - test if incidence of disease changes in a population over time

## Meta-analysis

- Combining results from a range of published studies
  
- Established methodology, not just literature review

# Qualitative designs

---

- If a concept or phenomenon needs to be understood because little research has been done on it, qualitative approach is preferred
- Qualitative research is exploratory and is useful when the researcher does not know the important variables to examine

This type of approach may be needed When:

- the topic is new,
- the topic has never been addressed with a certain sample or group of people, and
- existing theories do not apply with the particular sample or group under study

# Types of Qualitative design

---

- Qualitative research can be further classified in the following type.
  - I. **Phenomenology**:-a form of research in which the researcher attempts to understand how one or more individuals experience a phenomenon.
  - Eg:-we might interview 20 victims of tsunami tragedy
  - II. **Ethnography**:- this type of research focuses on describing the culture of a group of people.
  - A culture is the shared attributes, values, norms, practices, language, and material things of a group of people.
  - Eg:-the researcher might decide to go and live with the tribal in Andaman island and study the culture and the educational practices.

## Types of Qualitative design

---

- III. **Case study**: A form of qualitative research that is focused on providing a detailed account of one or more cases.
- IV. **Grounded theory**:- it is an inductive type of research, based or grounded in the observations of data from which it was developed;
  - ▣ it uses a variety of data sources, including quantitative data, review of records, interviews, observation and surveys
- V. **Historical research**:-it allows one to discuss past and present events in the context of the present condition, and allows one to reflect and provide possible answers to current issues and problems.
  - ▣ Eg:-the lending pattern of business in the 19th century.

# Mixed methods

---

- Mixed methods are becoming popular
- The problems addressed by social & health science researchers are complex, and the use of either quantitative or qualitative approaches by themselves is inadequate to address this complexity
- Mixed methods utilize the strengths of both qualitative & quantitative research
- The interdisciplinary nature of research, as well, contributes to the formation of research teams with individuals with diverse methodological interests and approaches

## Mixed methods cont...

---

A mixed methods design is useful when:

- either the quantitative or qualitative approach by itself is inadequate to best understand a research problem or
- the strength of both quantitative and qualitative research can provide the best understanding
- Eg. Researcher wanted to both generalize the findings to a population as well as develop a detailed view of the meaning of a phenomenon or concept for individuals
  - The researcher first explores generally to learn what variables to study and then
  - studies those variables with a large sample of individuals

# Mixed methods cont...

---

- Alternatively, researchers may first survey a large number of individuals and then
  - follow up with a few participants to obtain their specific language and voices about the topic
- In these situations, collecting both closed-ended quantitative data and open-ended qualitative data is advantageous
- The research project (mixed approach) will take extra time because of the need to collect and analyze both quantitative and qualitative data

# Setting

---

- ❑ In this part, basic information about the study area should be mentioned:
  - ❑ Health institutions
  - ❑ Schools
  - ❑ Community

# Population

---

## ❑ Who are your study subjects?

- ❑ Source population
- ❑ Study population
- ❑ Sampling unit (study unit)

## ❑ Inclusion and exclusion criteria

- ❑ Among those who are eligible to be the study participant, what are factors hindering to be included?
- ❑ Once mentioned in the inclusion criteria, then should not be repeated again in the exclusion part

# Assignment 4

---

- Choose appropriate design for your proposed research topic?
- Define the study setting (include necessary characteristics)
- Define the source and study population
- List the possible inclusion and exclusion criteria

# Session IV: Sampling and sample size

- Sampling techniques
- Sample size determination

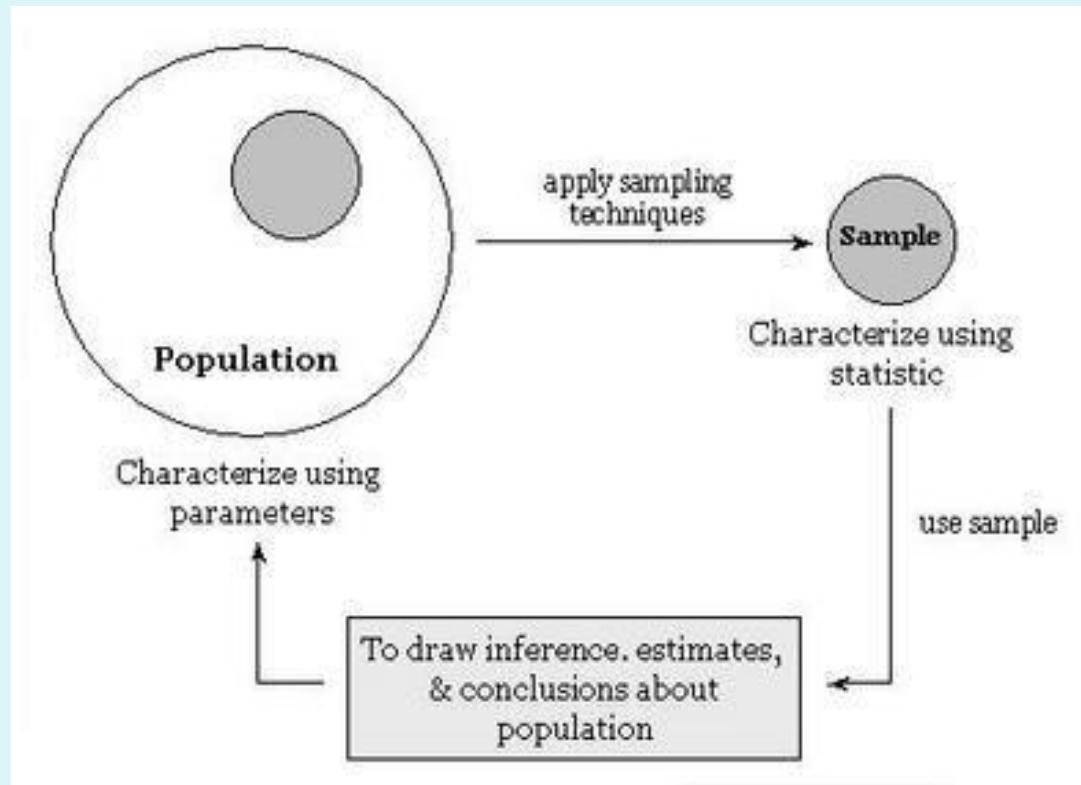
Aug 2018

# Sample

---

- In research terms a sample is a group of people, objects, or items that are taken from a larger population for measurement.
  
- The sample should be representative of the population to ensure that we can generalize the findings from the research sample to the population as a whole.

# Population and Samples



# Example

- House of representative
- Testing sauce while cooking
- Checking parts of a sack of Teff when you buy
- ...

# Concepts of Sampling

**Who do you want to generalize to?**

**What population can you get access to?**

**How can you get access to them?**

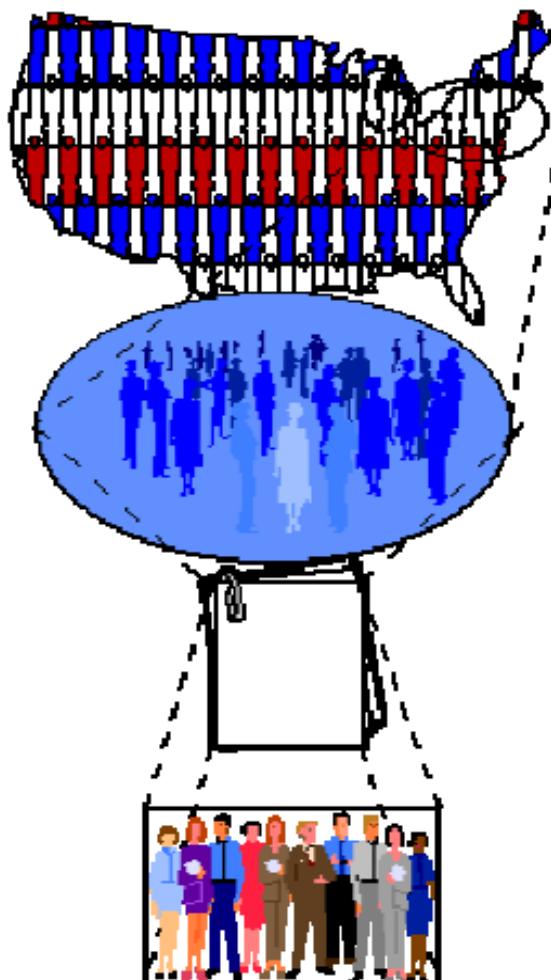
**Who Is In your study?**

**The Theoretical Population**

**The Study Population**

**The Sampling Frame**

**The Sample**



# Types of sampling

- A great deal of work has been done over the years in developing sampling methods that provide representative samples for the general population.
- There are different methods of sampling, broadly two:
  - ❖ Probability & non probability sampling

# Sampling methods

- Probability sampling

- Simple random sampling

- Systematic random sampling

- Stratified random sampling

- Cluster random sampling

- Multistage random sampling

- Non-probability sampling

- Purposive sampling

- Continence sampling

- Quota sampling

- Judgment sampling

- Snow ball sampling

# 1. Simple Random Sampling

- A sample size ‘n’ is drawn from a population ‘N’ in such a way that every possible element in the population has the same chance of being selected
- Assumption of the population
  - Homogeneity with respect to the variable of interest
  - If all members of a population were identical, the population is considered to be ***homogenous***
  - Availability of frame

## 2. Systematic random sampling

- Here are the steps you need to follow in order to achieve a **systematic random sample**:
  - ❖ number the units in the population from 1 to N
  - ❖ decide on the n (sample size) that you want or need
  - ❖  $k = N/n$  = the interval size
  - ❖ randomly select an integer between 1 to k
  - ❖ then take every  $k^{\text{th}}$  unit

# Example

- Select 20 samples from a total of 100 population using systematic random sampling technique?

**N = 100**

**want n = 20**

**N/n = 5**

**select a random number from 1-5:  
chose 4**

**start with #4 and take every 5th unit**

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

### 3. Stratified random sampling

- Divide the population into non-overlapping groups (i.e., *strata*)  $N_1, N_2, N_3, \dots N_i$ , such that  $N_1 + N_2 + N_3 + \dots + N_i = N$ . Then do a simple random sample depending on the type of allocation
- Elements **within** each strata are homogeneous, but are heterogeneous **across** strata.
- A **simple random** or a **systematic sample** is taken from each strata relative to the proportion of that stratum to each of the others

# Stratified random sampling...

- There are different sample allocation methods in order to select sample from each strata:

1. Proportional allocation: allocating sampling proportional to the total population of each strata using the formula:

$$n_i = \frac{n}{N} * N_i$$

- Where n=total sample size to be selected

- N=total population
- $N_i$  = total population of each strata
- $n_i$ =sample size from each strata

2. Equal allocation: allocating equal sample for each strata

## 4. Cluster Sampling

- A ***cluster sample*** is a simple random sample of groups or clusters of elements (vs. a simple random sample of individual objects).
- This method is useful when it is difficult or costly to develop a complete list of the population members or when the population elements are widely dispersed geographically.
- Cluster sampling may increase sampling error due to similarities among cluster members.

## 5. Multi-stage Sample Designs

- Many surveys use complex sample designs that combine several of the above elements in a multi-stage sampling framework
- Sometimes the population is too large and scattered for it to be practical to make a list of the entire population from which to draw a SRS
- Suppose that each unit in the population can be divided into a number of smaller units, or subunits

# Non-probability sampling

- Convenience sampling
  - *Drawn at the convenience of the researcher. Common in exploratory research. Does not lead to any conclusion.*
- Judgmental sampling
  - *Sampling based on some judgment, gut-feelings or experience of the researcher. If inference drawing is not necessary, these samples are quite useful.*
- Quota sampling
  - *An extension of judgmental sampling. It is something like a two-stage judgmental sampling. Quite difficult to draw.*
- Snowball sampling
  - *Used in studies involving respondents who are rare to find. To start with, the researcher compiles a short list of sample units from various sources. Each of these respondents are contacted to provide names of other probable respondents.*

# Qualitative Sampling Techniques

- Interested in samples of participants who can share their interpretation of the experience with others
- Goal is understanding the meaning of the participants' experience
- Typically not interested in generalizing their results
- Typically do not use probability sampling

# Sample Size Determination

- How Big is Big Enough?
- Generally the larger the better, but that takes more time and money.
- Answer depends on:
  - How different or dispersed the population is.
  - Desired level of confidence.
  - Desired degree of accuracy.
  - Desired margin of error

## Sample size ...

- Which variables should be included in sample size calculation?
  - It should relate to the study's primary outcome variable
  - If the study have secondary outcome variables which are considered important, the sample size should also be sufficient for the analysis of these variables.

# Sample size determination

- There are different approaches to determine size of the sample
  - Rules of thumb approach
  - Confidence interval approach
  - Hypothesis testing approach

# Confidence interval approach

- Given confidence interval

$$\text{mean(proportion)} \pm z_{\frac{\alpha}{2}} s.e$$

- Hence the absolute precision denoted by  $d$  is given as

$$d = z_{\frac{\alpha}{2}} s.e$$

- Where  $s.e$  is the standard error of the estimator of the parameter of interest.

# Estimating a single population mean

$$\begin{aligned}d &= Z_{\frac{\alpha}{2}} SE \\&= Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ \Rightarrow \sqrt{n} &= Z_{\frac{\alpha}{2}} \frac{\sigma}{d} \\ \Rightarrow n &= Z_{\frac{\alpha}{2}}^2 \frac{\sigma^2}{d^2}\end{aligned}$$

# Single population proportion

- Let  $p$  denotes proportion of success, then

$$\begin{aligned}d &= Z_{\frac{\alpha}{2}} SE \\&= Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}} \\\Rightarrow \sqrt{n} &= Z_{\frac{\alpha}{2}} \frac{\sqrt{P(1-P)}}{d} \\\Rightarrow n &= Z_{\frac{\alpha}{2}}^2 \frac{P(1-P)}{d^2}\end{aligned}$$

# Single ...

- The possible source of this proportion are:
  - ✓ From the results of a previous study
  - ✓ From a pilot study
  - ✓ Judgment of the researcher.
  - ✓ Simply taking 50%
- Maximum acceptable difference (d):
  - It takes a value between 1% and 5%
  - Commonly  $d=5\%$
- Desired confidence level
  - At 95% confidence level  $Z_{\alpha/2}=1.96$

# Example

- It was desired to estimate proportion of anaemic children in a certain preparatory school. In a similar study at another school a proportion of 30 % was detected. Compute the minimal sample size required at a confidence limit of 95% and accepting a difference of up to 4% of the true population.

# Example

- In previous studies, percentage of hypertensives among Diabetics was 70% and among non diabetics was 40% in a certain community. A researcher wants to perform a comparative study for hypertension among diabetics and non-diabetics at a confidence limit 95% and power 80%, What is the minimal sample to be taken from each group with 4% accepted difference of true value?

# Example

- A study is to be performed to determine a certain parameter in a community. From a previous study a sd of 46 was obtained. If a sample error of up to 4 is to be accepted. How many subjects should be included in this study at 99% level of confidence?

# Variables

---

- Dependent/independent variables
- Operationalizing variables (making them measurable)

# Operational definitions

- Many variables can easily be measured
- For some variables it is sometimes not possible to find meaningful categories unless the variables are made operational with one or more indicators
- Operationalizing variables means that you make them measurable

# Operational ...

- Ex 1- determining level of knowledge
- There is a need to develop a series of questions to assess knowledge
- The answer to these questions form an indicator of someone's knowledge on the issue, which can then be categorized
- E.g if 10 questions were asked, you might decide that the knowledge of those with:
  - 0 to 3 correct answers is poor
  - 4 to 6 correct answers is reasonable, and
  - 7 to 10 correct answers is good

# Operational ...

- Ex2- Nutritional status of <5 children
- Cannot be measured directly
- You would need to choose appropriate indicators
- Widely used indicators for nutritional status include:
  - ▣ Weight for age
  - ▣ Weight for height
  - ▣ Height for age
- Weight for age (Based on standard growth curve)
  - ▣ Well nourished – if they are above 80% of the standard
  - ▣ Moderately malnourished – if they are between 60% and 80%
  - ▣ Severely malnourished – if they are below 60%

# Assignment 5

---

- For your proposed topic of research:
  - Define the target population
  - Define the study population
  - What is the appropriate sampling method
  - Determine sufficient sample size

# Session V: Data collection

- Techniques of data collection
- Tools of data collection
- Questionnaire design
- Bias, validity, and reliability

Aug 2018

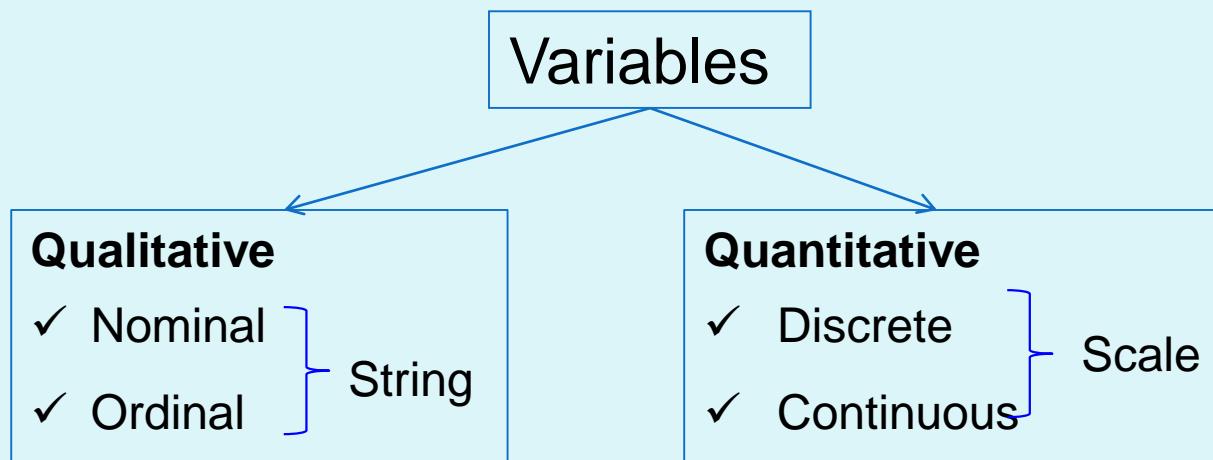
# Learning Objectives

At the end of this session the student will be able to:

- Understand the different techniques and tools of data collection
  
- Use appropriate data collection techniques when planning and conducting research

# Introduction (1)

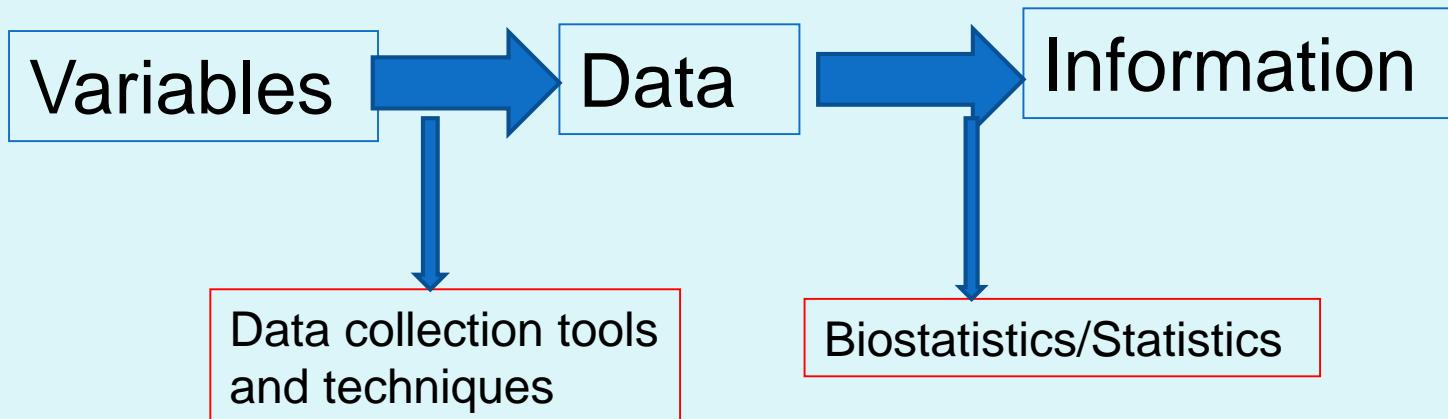
- To collect data -→ develop tools of data collection



- To analysis data-→statistical packages

## Introduction (2)

- Methods of data Analysis depends on the tpyes of variable, objective and sample size

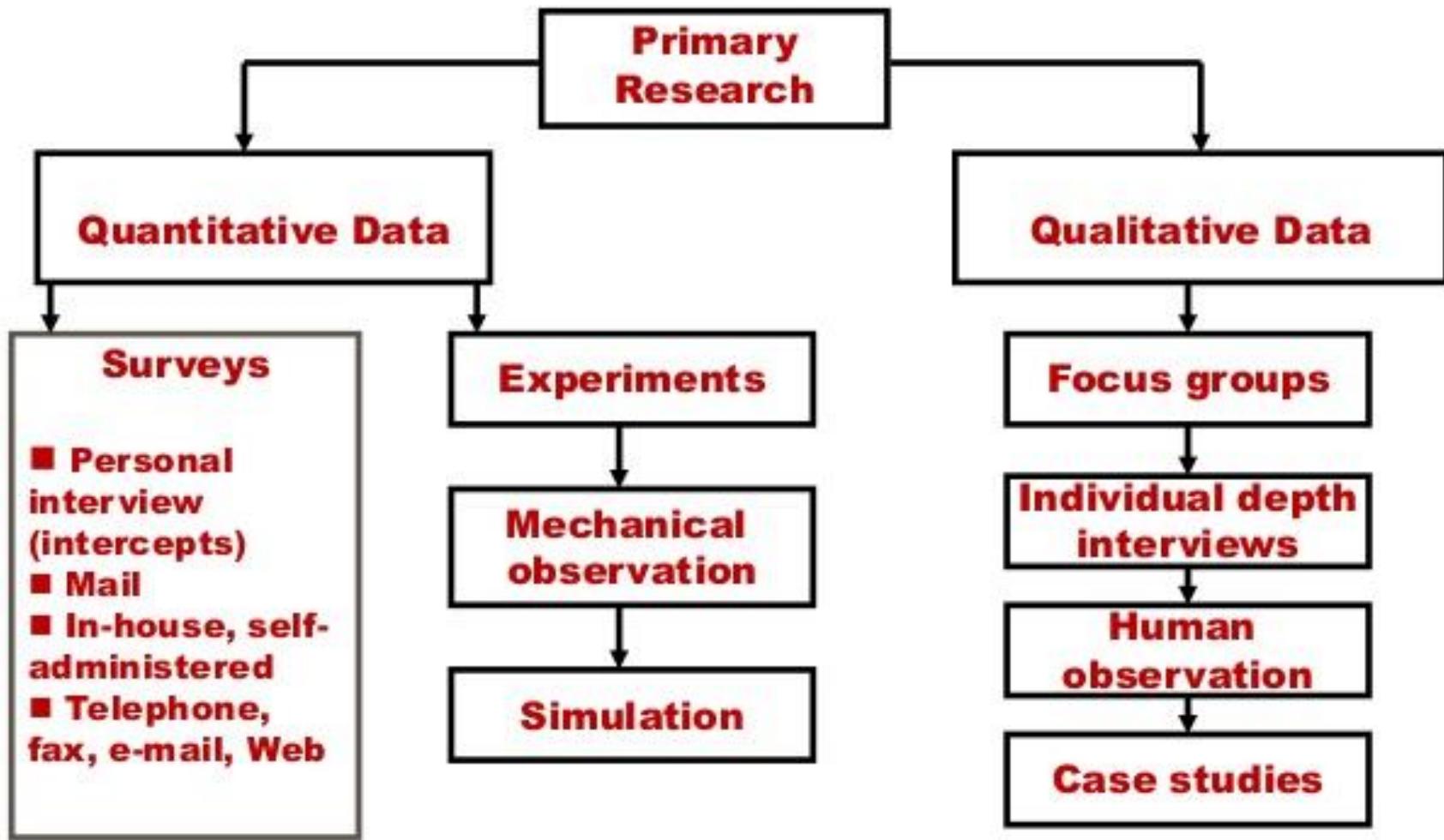


- ✓ Data entry/cleaning
- ✓ Descriptive statistics
- ✓ Inferential statistics

# Introduction (3)

- Statistics is the tool which converts data to information and can be:
  - Descriptive statistics
    - Proportion/ratio/rate
    - Measures of central tendency
    - Measures of dispersion
  - Inferential statistics
    - Estimation
    - Hypothesis
    - Statistical modeling

# Primary Research methods



# Overview of data collection techniques

## Quantitative Data collection

- Using available information
- Observing
- Interviewing
- Administering written questionnaires
  
- Projective techniques, mapping, scaling

## Qualitative data collection

- Focus group discussions
- In-depth interview
- Observation
- Key informant interview

# 1. Available information

- For ex- analysis of the information routinely collected by health facilities can be very useful for identifying problems
- Other sources of available data – newspapers, published case histories etc
- **Advantage of available data** – collection is inexpensive
- **Disadvantage of existing data:**
  - It is some times difficult to gain access to records or reports
  - Data may not always be complete and precise enough, or too disorganized

## 2. Observation

- Observation is a technique that involves systematically selecting, watching and recording behavior and characteristics of living beings, objects or phenomena
- Different ways:
  - participant observation – observer takes part in the situation he or she observes
  - Non-participant observation – observer watches the situation, openly or concealed, but does not participate

# Observation cont...

- Observation can give additional, more accurate information on behavior of people than interviews or questionnaires
- Observations of human behavior can form any type of study, but as they are time consuming they are most often used in small scale studies
- Observations can also be made on objects
  - the presence or absence of latrines and the state of cleanliness may be observed

### 3. Interviewing

- It involves oral questioning of respondents, either individually or as a group
- Answers can be recorded by writing them down or by tape-recording the responses, or by a combination of them
- Interviews can be conducted with varying degree of flexibility (high degree of flexibility Vs low degree of flexibility)

**High degree of flexibility:** Usually used when studying sensitive issues or when the researcher has little understanding of the problem

- Is frequently applied in exploratory studies

# Interviewing...

- When studying sensitive issues (e.g teenage pregnancy) the investigator may use a list of topics rather than fixed questions
- The sequence of topics should be determined by the flow of discussion
- It is often possible to come back to a topic discussed earlier in a later stage of the interview

## 4. Self-administered questionnaire

- Written questions are presented that are to be answered by the respondents in written form
- A written questionnaire can be administered in different ways, such as by:
  - Sending questionnaires by mail
  - Gathering all or part of the respondents in one place at one time, giving oral or written instructions, and letting them fill out the questionnaires
  - Hand-delivering questionnaires to respondents and collecting them later

# Qualitative study

- Observation
- FGDs allow a group of 8-12 informants to freely discuss a certain subject with the guidance of a facilitator or reporter
- In-depth interview
- Key informant interview
- ...

# Differences between data collection techniques and data collection tools

Data collection techniques	Data collection tools
Using available information	Checklist; data compilation forms
Observation	Eyes and other senses, pen/paper, watch, scales, microscope, etc..
Interviewing	Interview guide, checklist, questionnaire, tape recorder
Administering written questionnaire	Questionnaire

# **QUESTIONNAIRE DESIGN**

# Questionnaire Design

- The 2 important procedures at the outset of constructing a questionnaire are planning & piloting
- In the planning of the questionnaire, it is important to:
  - list the topics of interest in relation to the aims of the study,
  - collate appropriate and tested questions and scales,
  - list additional items and response formats that need to be developed, and finally relate the questions back to the survey aims

# Piloting

- The ideas and topics should be tested on colleagues and then pre-piloted with a small number of in-depth interviews
- The investigator should hold meetings with experts in the field and group discussions with members of the target group in order to ensure validity of the coverage
- If the questionnaire contains new, previously untested items, then they shall be tested face to face on a sample of people from the target population (about 30-50, depending on the complexity of the items)

# Issues to be addressed in the pilot study

- Is each question measuring what it is intending to measure?
- Is the wording understood by all respondents, and is the understanding (meaning) similar for all respondents?
- Are the instructions on self-administered questionnaires understood by all respondents?

# Questionnaire layout

- The questionnaire should be visually easy to read and comprehend
- Lower case letters should be used for texts  
It is customary for the first few lines of a questionnaire to include:
  - the label ‘confidential’,
  - the respondent’s serial (identification) number (to preserve anonymity),
  - the title of the study and
  - a brief introduction

# Question numbering and topic ordering

- Questions must be numbered (1,2 etc), and sub-questions clearly labeled (e.g 1a, 1b etc)
- A question and its responses should never be split over two pages, as this can lead to confusion
- The order of questions is important
- Questions should not skip backward and forward between topics

# The covering letter

- The covering letter should:
  - be written on the organization's headed paper,
  - include the name and address of the sample member and the identification (serial) number, and
  - address the recipient by name

The letter should:

- explain how the person's name was obtained,
- outline the study aims and benefits (concisely),
- guarantee confidentiality
- be signed in blue ink

# Question forms

- Question form refers to the format of the question (closed or open-ended), and type of measuring instrument
- The comprehensiveness of response choices for closed questions is also important
- Response choices to questions can be left open or they can be closed or ‘pre-coded’
- Closed questions can be:
  - dichotomized (e.g yes/no response choices),
  - multiple response (no restrictions on the number of responses that can be ticked) or
  - scaled (with one response code per response frame permitted)

# Question forms cont...

- **Structured questionnaires** - involve the use of fixed questions or scales which are presented to respondents in the same way, with no variation in question wording and with closed ques (pre-coded response choices)
- Some structured questionnaires will also include open-ended questions, to enable respondents to reply in their own words

# Question forms cont...

- **Semi-structured interviews** include fixed questions but with no, or few, response codes, and are used flexibly, often in no fixed order, to enable respondents to raise other relevant issues not covered by the interview schedule
- **Unstructured interviews-** are comprised of a checklist of topics, rather than fixed questions, and there are no pre-codes
- The more structured approach is only suitable for topics where sufficient knowledge exists for largely pre-coded response formats to be developed

# Data Quality Control Issues

Describe/provide:

- Operational definitions of crucial concepts
- Selection and training of field workers/research staff
- Field testing the research methods and tools
- Supervision and quality control
- Pretest** – usually refers to a small-scale trial of particular research components

# Bias, Validity, and Reliability

- In **research**, **bias** occurs when systematic error is introduced into sampling or testing by selecting or encouraging one outcome or answer over others
- Validity, in general, is an indication of how sound your research is
- **Validity** in data collection means that your findings truly represent the phenomenon you are claiming to measure
- **Reliability** is also a way of assessing the quality of the measurement procedure used to collect data in research
- In order for the results from a study to be considered valid, the measurement procedure must first be reliable

## Exercises: Design template for the following questions

S. No	Question	Category /Answer	Skip to
1	ID	_____	
2	Field of study	1.Internal medicine 3. Pediatrics 5. Gynecology 2. radiology 4. surgery 6. Other	
3	Sex	1. Male    2. Female	
4	Age at first formal school	1.Age in years _____	
5	Family residence	1. Urban    2. rural 3. semi-urban	
6	Current marital status	1. Never married    2. Married	
7	Smoking status	1. Yes    2. No	If No for Q 7, go Q10.
8	Frequency of smoking	1. Daily    2. weekly    3. monthly	
9	Number per day	-----	
10	Have you ever participated in research before ?	1.yes 2. No	If No for Q 8, go to Q12.
11	Type of research	1. Community based 3. Clinical trial 5. other , specify 2. Institute based 4. Survey	
12	How do you level the added knowledge of research course for your career?	1.Excellent    2. Good 3. little    4. Not at all	

# Assignment

---

- Prepare appropriate data collection tool for your research topic?

# Session VII

## Data Analysis and Presentation

Aug 2018

# Outline

---

- Data analysis
- Variable definition
- Data entry
- Data management
- Descriptive data analysis
- Inferential data analysis

# Learning Objectives

---

- Compute descriptive statistics
- Present outputs in tables and graphs
- Define inferential statistics
- Describe the model formulation for the statistical model
- State assumptions of the model
- Apply appropriate the model for the given data
- Interpret the output and present the findings

# Data management

---

- Data processing refers to:
  - data entry into a computer
  - data checks and correction
  - Variable coding
  - Data cleaning
- The aim of this process is to produce a relatively 'clean' data set

# Data management cont...

---

## **Data entry**

- Data entry concerns the transfer of data from a questionnaire to a computer file.
- It is converting data into a form that can be read and manipulated by computers used in quantitative data analysis.
- Before analysis, data must be checked for errors, information that needs coding must be coded and missing values must be dealt with

# Data management cont...

---

## Data coding

- In general computers are at their best with numbers.
- Some statistical packages cannot analyze alphabetic codes, some cannot understand open ended responses,
- You must translate variables through ***coding***
- Coding is assigning a separate (non-overlapping) numerical code for separate answers and missing values.
- E.g. Instead of using 'Male' and 'Female' for

# Data management cont...

---

- Missing values occur when measurements were not taken, or respondents did not answer questions.
- In general, missing values should not be entered as a '***blank***', ***because some statistical packages interpret blanks as zero.***
- ***Ideally, a*** code should be chosen to denote a missing value (e.g. a code '9' or '99' or '999')
- After the analysis is finished, decoding back to original variables is required when

# Data management cont...

---

## **Data Cleaning**

- Once data have been gathered, they need to be entered into a computer data file and checked for errors.
  
- No matter how carefully the data have been entered some errors are inevitable.
  
- The aim of this process is to produce a clean set of data for statistical analysis.

# Data Analysis and Presentation

---

- Data Analysts are in high demand in all areas of employment
- The goal of data analysis is to discover useful information, suggesting conclusions, and supporting decision-making
- Sound policies rest on good information
- It has multiple approaches ranging from descriptive to inferential analysis
- Descriptive analysis includes
  - Table
  - Diagram
  - Summary measures

# Inferential statistics

---

- The central activity in statistics is inference
- The goal is to make conclusions about a phenomenon based on observed data
- In this way, statistical inference may help the researchers in suggesting or verifying scientific hypotheses, or decision makers in improving their decisions
- There are many Statistical modes of performing inference
- Good statistical inference never strays very far from the data

# Choosing Statistical models

---

- The choice of statistical models for the specific data depends on different properties:
  - Objective of the study
  - Study Design
  - Nature of the variable
  - Distribution of the variable
  - The nature of the data
  - The number of group we want to compare
  - Sample size

# Comparing two groups

---

- ❑ The outcome variable is continuous and we want to compare among two groups
  - ❑ Example: Weight among gender
- t-test (3 types of t-test)
  - ❑ One sample t-test
  - ❑ Independent sample t-test
  - ❑ Paired sample t-test

# Comparing more than two groups

---

- Two compare among more than two groups
  - ANOVA
  - Example: Weight among mothers age group (<20 years, 20-30, and >30 years)
- Assumption
  - The outcome variable is continuous
  - The error term is assumed to be normally distributed
  - Variance is assumed to be constant
  - Factor variable is categorical with more than two groups

# Hypothesis to be tested

---

- The hypothesis to be tested are
  - $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ 
    - Versus
  - $H_0: \mu_i \neq \mu_j, i \neq j$
  - The test statistics is F
  - The tabulated value test of F at  $\alpha$  level of significant and df1 and df2 degree of freedom will be obtained from the table where
  - df1= is degree of freedom of the numerator
  - df2= is degree of freedom of the denominator

# Hypothesis test

---

- The calculated value of the test statistics can be obtained from the ANOVA table

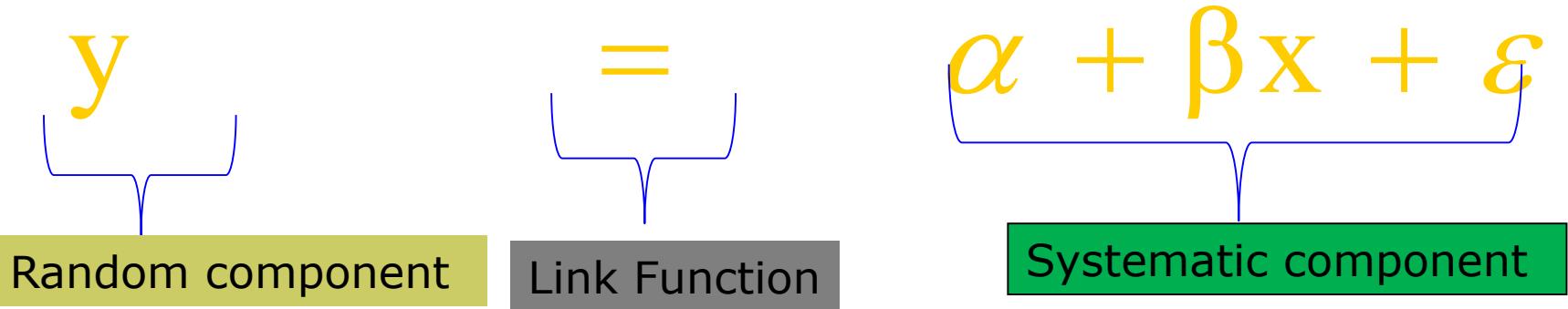
<b>Source</b>	<b>df</b>	<b>SS</b>	<b>MS</b>	<b>F</b>	<b>P-value</b>
Between	K-1	SSR	$MSR = SSR/K-1$	$MSR/MS_E$	
Within	N-k	SSE	$MSE = SSE/N-K$		
Total	N-1	SST			

- Then if  $F > F_{tab}$  then reject the null hypothesis
- Otherwise, accept the alternative hypothesis

# Statistical model

## ❑ Components of statistical model

- Random component
- Systematic component
- Link function



# Regression

---

- Outcome variable: Continuous
- Research question: Prediction of one variable by another
  - Simple Linear Regression

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Multiple linear regression

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_l x_l + \varepsilon_{ij}$$

- Assumptions of the model
  - Linearity
  - Normality:
  - Independent: the observations are assumed to be uncorrelated
  - Homoscedasticity

# Regression

---

- When the observations are correlated:
  - Linear mixed model (**LMM**)

$$Y_i = \beta_i X_i + Z_i b_i + \varepsilon_i$$

- Where  $\beta$  is the parameters of the fixed effect
- $X_i$  is design matrix of fixed effect covariates variables
- $Z_i$  is design matrix for random effect covariates
- $b_i$  is the vector of parameters for the random effects
- And  $\varepsilon_i$  is random effort term

# Hypothesis test

---

- In this modeling approach, we test the following hypothesis

$$H_0 : \beta_i = 0 \text{ Versus } H_a : \beta_i \neq 0$$

- We use 95% CI and check if the interval includes zero
- We can also use p-value if it is lower than 0.05
- The goodness of fit test can be tested using multiple determination ( $R^2$ )

# Logistic regression

---

- Outcome variable: categorical
- Research question: Prediction
- Statistical model: Logistic regression

- Simple logistic regression

$$\ln\left(\frac{P}{1-P}\right) = \text{logit}(p) = \alpha + \beta x$$

- Multiple logistic regression

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i$$

# Logistic Regression

---

- ❑ Hypothesis to be tested

$$H_0 : OR_i = 1 \text{ Versus } H_a : OR_i \neq 1$$

- ❑ We use 95% CI and check if the interval includes one
- ❑ We can also use p-value if it is lower than 0.05
- ❑ The goodness of fit test can be tested using Hosmer-Lemshow test
- ❑ We can also use the **pseudo-R<sup>2</sup>**

# Extension of logistic regression

---

- When the assumption of ordinary logistic regression didn't work, we use its extensions

- Multinomial logistic regression
- Ordinal logistic regression
- Mixed effect logistic regression
- Generalized linear mixed model
- Generalized estimating Equation

Read about the  
two models?

# Poisson regression

---

- Outcome variable: Count
- Research question: Prediction
  - Statistical model: Poisson Regression
- A random variable  $Y$  is said to have a Poisson distribution with parameter  $\mu$  if it takes integer values  $y = 0, 1, 2, \dots$  with probability

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}, \text{ for } \mu > 0$$

- Assumption: Mean=variance

# Poisson regression

---

- The model is formulated as

$$\log(\mu / t) = \alpha + \beta x$$

- Check Overdispersion
  - Mean and variance are not the same ( $\text{Var}(Y) = \phi\mu$  where  $\phi$  is a scale parameter)
- Overall goodness-of-fit statistics of the model are the same as for any GLM:
  - Pearson chi-square statistic
  - Deviance
  - Likelihood ratio test, and statistic

## Hypothesis test

---

- The hypothesis to be tested are

$$H_0 : RR_i = 1 \text{ Versus } H_a : RR_i \neq 1$$

- We use 95% CI and check if the interval includes one
  - Wald statistics
  - Likelihood ratio tests
  - Score tests
- We can also use p-value if it is lower than 0.05

## Extensions of Poisson regression

---

- When the assumption of Poisson regression didn't work, then we use:
  - Zero-inflated Poisson regression
  - Negative binomial Poisson regression
  - Zero-inflated negative binomial Poisson regression
  - Hurdle regression model
  - Mixed effect Poisson regression

# Survival Analysis

---

- Outcome variable: time to event
- Statistical approach: survival analysis
  - Notation:
    - $T \equiv$  survival time of a selected individual
    - $t \equiv$  a specific point in time.
    - $S(t) =$  Survival Function
    - $h(t) \equiv$  instantaneous failure rate among survivors at time  $t$  (*also called hazard function*)

# Survival analysis

---

- The survival function can be formulated as

$$f(t) = \lim_{it(\Delta t \rightarrow 0)} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Then survival probability is

$$\hat{S}(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(u) du$$

- Life table is the probability calculated from the survival function
- It takes censoring in to account to compute probabilities of the life table

## Kaplan–Meier curve

---

- ❑ The probability of surviving time  $t$ ,  $S(t)$ .

- ❑ Let the observed times until event of the  $n$  sample members be

$$t_1 \leq t_2 \leq \dots \leq t_n$$

- ❑ Corresponding to each  $t_i$  is  $n_i$ , the number at risk just prior to time  $t_i$ , and  $d_i$ , the number of events at time  $t_i$

- ❑ Then the Kaplan–Meier estimator

$$\hat{S}(t) = \prod_{i|t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right),$$

- ❑ It assumes that censoring is independent of developing the event

# The Cox-regression model

---

- A baseline hazard is the hazard when all covariates are 0 which is denoted by  $h_0(t)$
- The hazard at time  $t$  of covariate  $i$  is defined by  $h_i(t)$

$$h_i(t) = h_0(t) \exp(\beta_1 x_1)$$

- Is univariate cox regression model for covariable  $x$

$$\log h_i(t) = \log h_0(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

- Is multiple cox-proportional hazards model

$$h_i(t) = h_0(t) e^{\beta_1 x_{i1} + \dots + \beta_k x_{ik}}$$

# The Cox-regression model

---

- Hypothesis to be tested

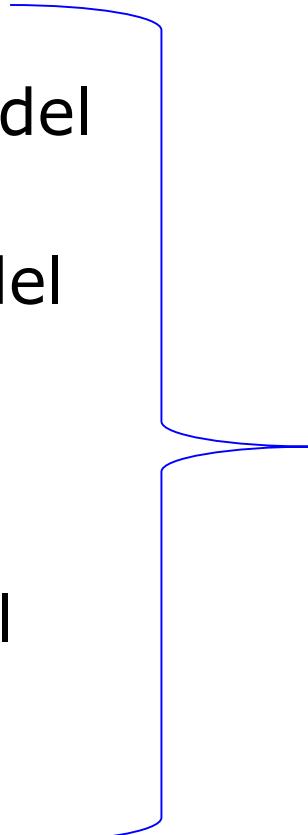
$$H_0 : HR_i = 1 \text{ Versus } H_a : HR_i \neq 1$$

- We use 95% CI and check if the interval includes one
  - Wald statistics
  - Likelihood ratio tests
  - Score tests
- We can also use p-value if it is lower than 0.05

# Extensions of Survival analysis

---

- Recurrent event model
- Competing risk model
- Multistage model
- Extended cox-model
- Joint modeling



Read these models?

# Data analysis using SPSS



1. Descriptive data analysis
2. Inferential data analysis

# Analyze: Descriptive Data Analysis

- Assuming that the data is ready for analysis, we start exploring using descriptive analysis

The screenshot shows the SPSS interface with the 'Analyze' menu highlighted in yellow. A black callout arrow points from the text 'Analyze ---->' to the 'Analyze' menu item. To the right of the menu, a list of statistical procedures is displayed. The data view window in the background shows a table with columns 'idno' and 'TotPreg'.

	idno	TotPreg
1	7.00	5.0
2	47.00	4.0
3	49.00	3.0
4	52.00	5.0
5	56.00	5.0
6	58.00	4.0
7	64.00	3.0
8	67.00	6.0
9	68.00	3.0
10	70.00	5.0
11	71.00	8.0
12	75.00	1.0
13	76.00	6.0
14	78.00	1.0
15	79.00	1.0
16	84.00	2.0
17	85.00	3.0
18	86.00	5.0
19	91.00	1.0
20	92.00	1.0

Tadesse A.

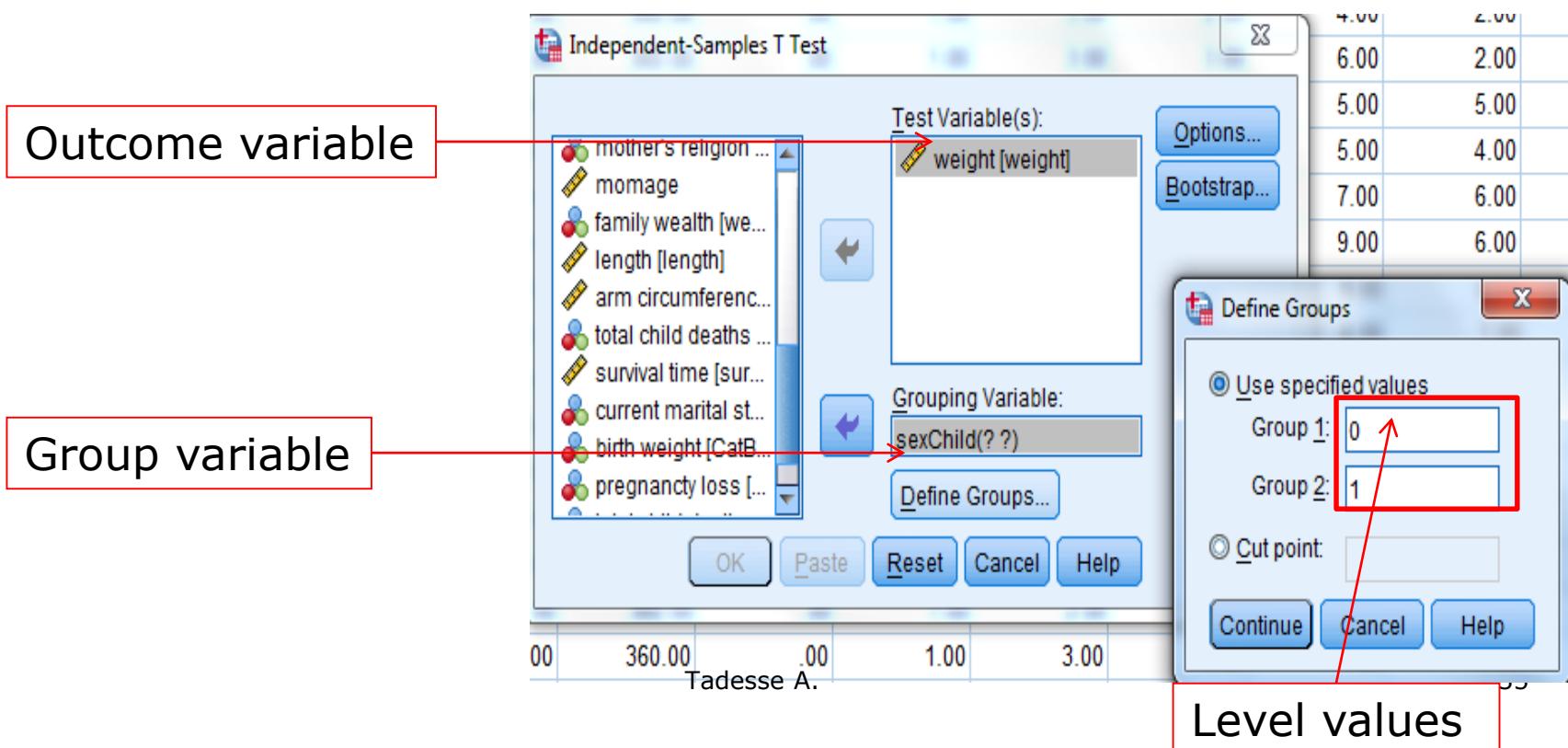
File Edit View Data Transform Analyze Direct Marketing Graphs

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...

34

# Independent sample t-test

- To compare the outcome variable among two levels of factor variable
  - Eg. Comparing birth weight among sex of the child



# Outputs

---

Group Statistics

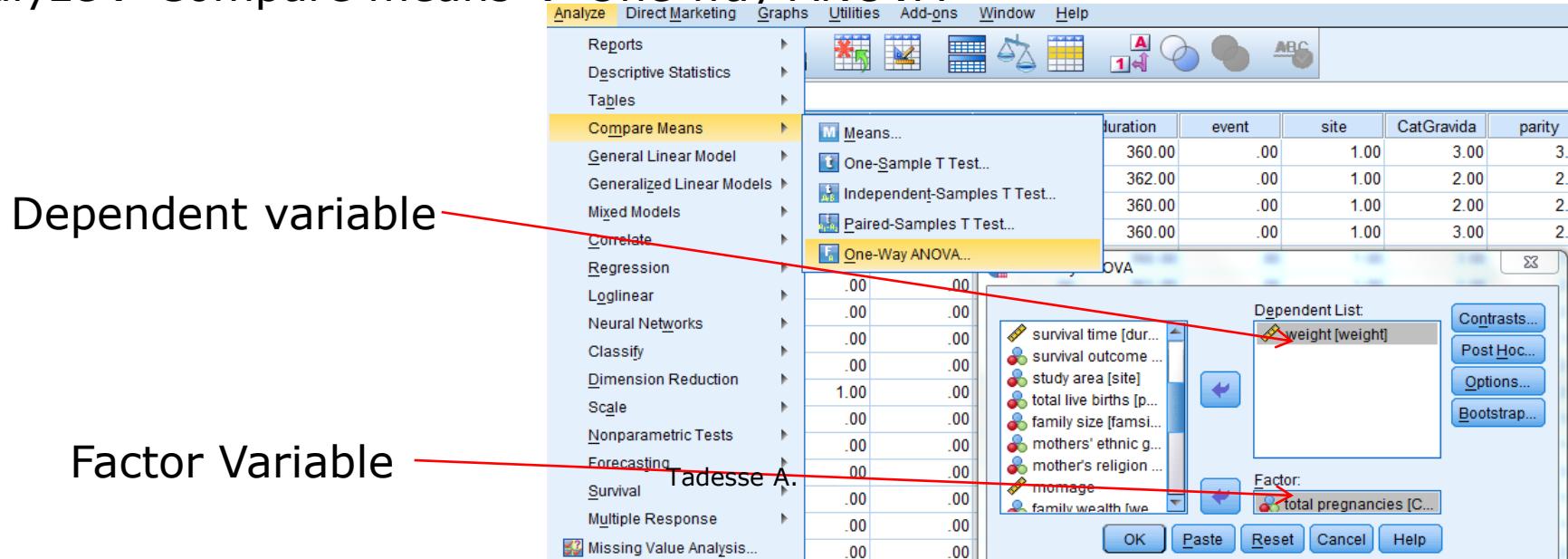
	sex	N	Mean	Std. Deviation	Std. Error Mean
weight	.00	3861	3055.3227	503.28204	8.09956
	1.00	4012	3169.7241	521.84525	8.23875

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
weight	Equal variances assumed	-9.895	7871	.000	-114.40136	11.56138	-137.06473	-91.73800
	Equal variances not assumed	-9.902	7870.964	.000	-114.40136	11.55335	-137.04900	-91.75373

# Analysis of Variance (ANOVA)

- ❑ If the factor/group variable has more than two levels, t-test is not useful to compare the outcome variable between the groups
- ❑ Analysis of Variance (ANOVA) is used to compare between groups if more than two

Analyze → Compare means → One way ANOVA



# Outputs

Descriptives								
weight	N	Mean	Std. Deviation	Std. Error	95% CI for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	1746	2969.7892	508.76248	12.17567	2945.9088	2993.6697	1400.00	5600.00
2-4	3522	3152.0409	504.44631	8.50003	3135.3754	3168.7064	1180.00	5800.00
>4	2605	3158.0783	519.07858	10.17019	3138.1358	3178.0208	1280.00	5500.00
Total	7873	3113.6205	515.97277	5.81509	3102.2213	3125.0196	1180.00	5800.00

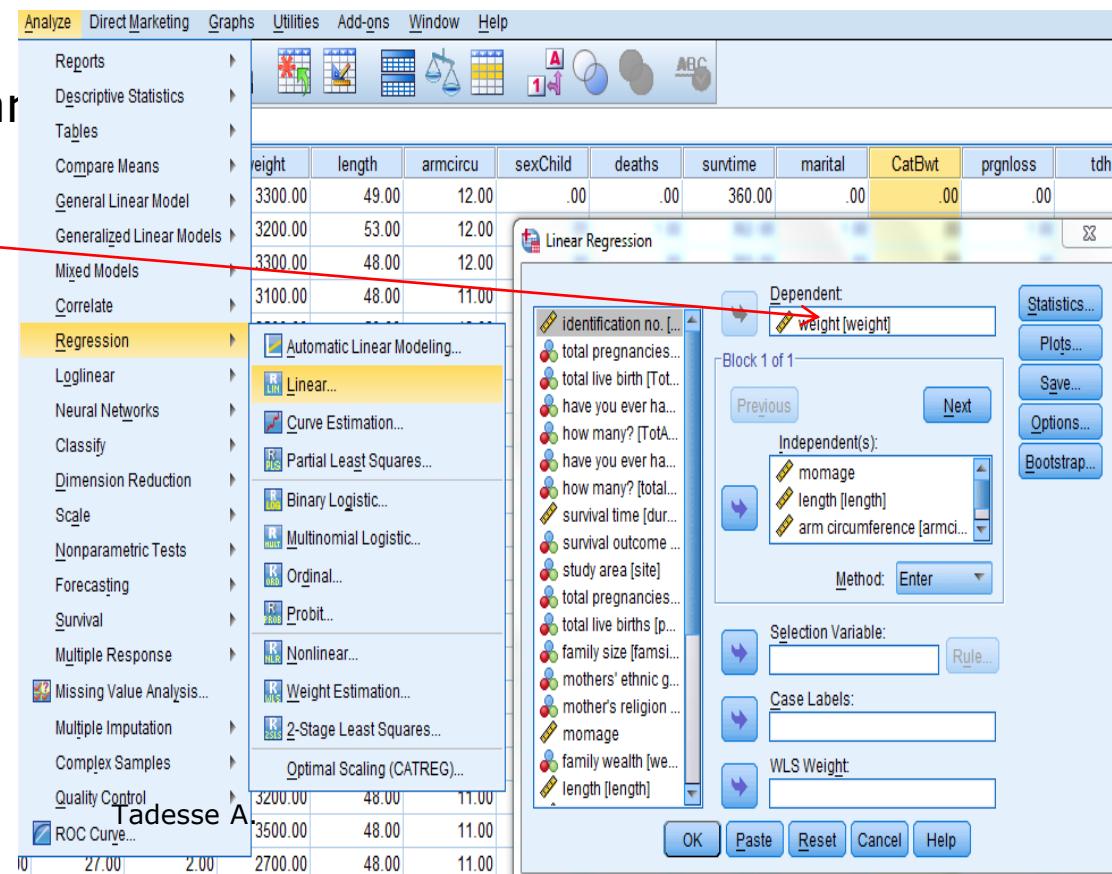
ANOVA					
weight	Sources	Sum of Squares	df	Mean Square	F
Between Groups	46467949.405	2	23233974.702	89.227	.000
Within Groups	2049278048.57	7870	260391.112		
Total	2095745997.97	7872			

# Linear regression

- We may be interested in predicting the effect of independent variables on the outcome variable which is continuous

Analysis → regression → linear

Dependent variable



# Outputs

Model		Coefficients <sup>a</sup>					
		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta			
1	(Constant)	2868.003	25.521			112.380	.000
	momage	-.960	1.052	-.012	-.913		.361
	family size	38.388	3.120	.158	12.305		.000
	sex	114.982	11.427	.111	10.062		.000

a. Dependent Variable: weight

# Logistic regression

---

- Now, the outcome variable is categorical
- If the categorical outcome variable is binary (only 2 possible values), then binary logistic regression is fitted
- If the outcome variable has more than two levels of nominal, then multinomial logistic regression
- If the outcome variable has ordered levels, then ordinal logistic regression is fitted

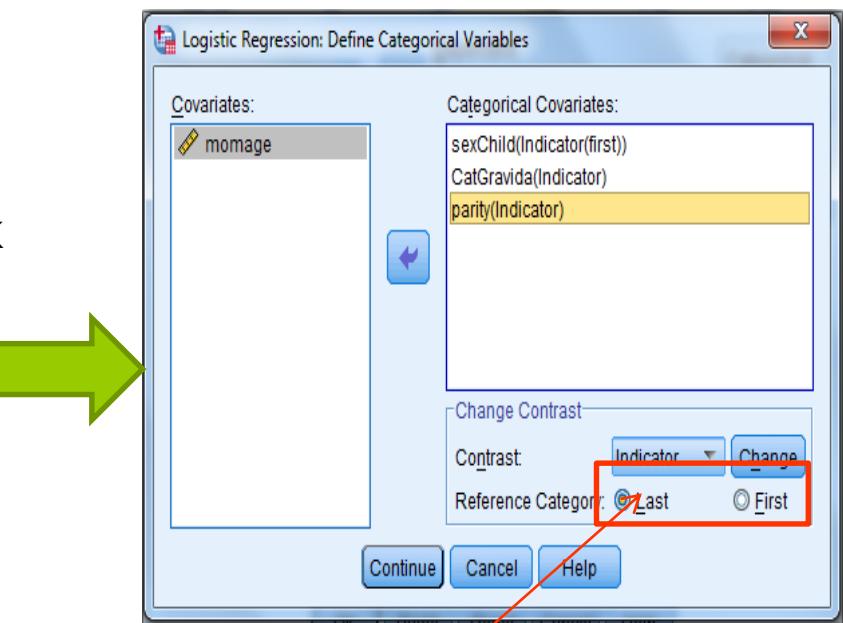
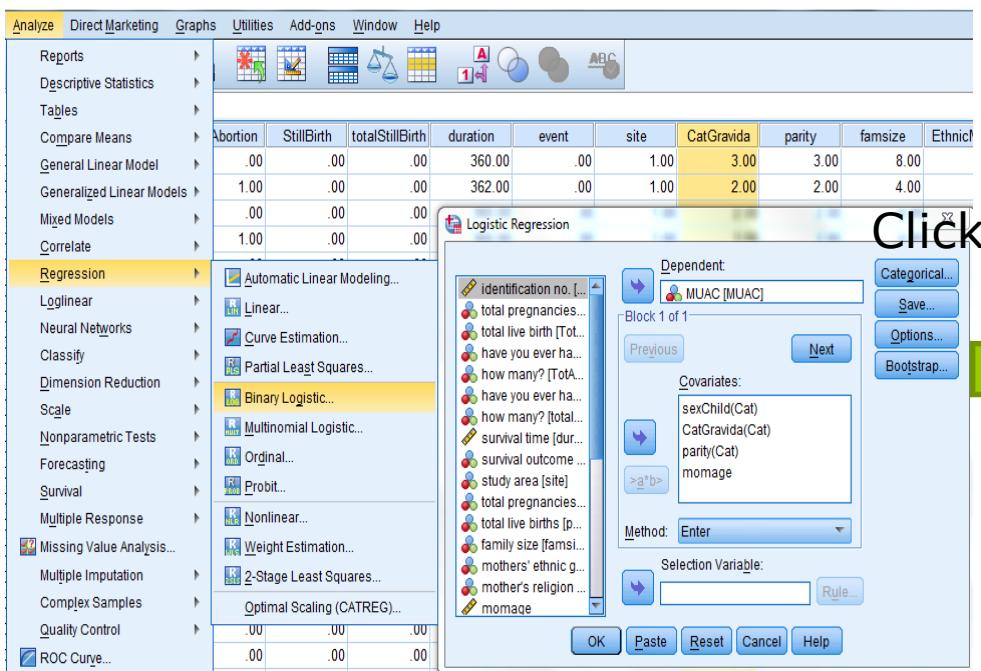
# Binary logistic regression

---

- **Dependent variable:** Birth weight (MUAC, normal=1, underweight =0)
- **Explanatory variable:** Age of mother, sex of the child, gravidity, parity, marital status, and so on
  - **Univariate logistic regression:** Fitting logistic regression for each explanatory variable
  - 
  - Odds ratio, 95% CI and p-value will be computed
  - A variable will be taken to multivariable logistic regression if p-value is less than 20%

# Binary logistic regression

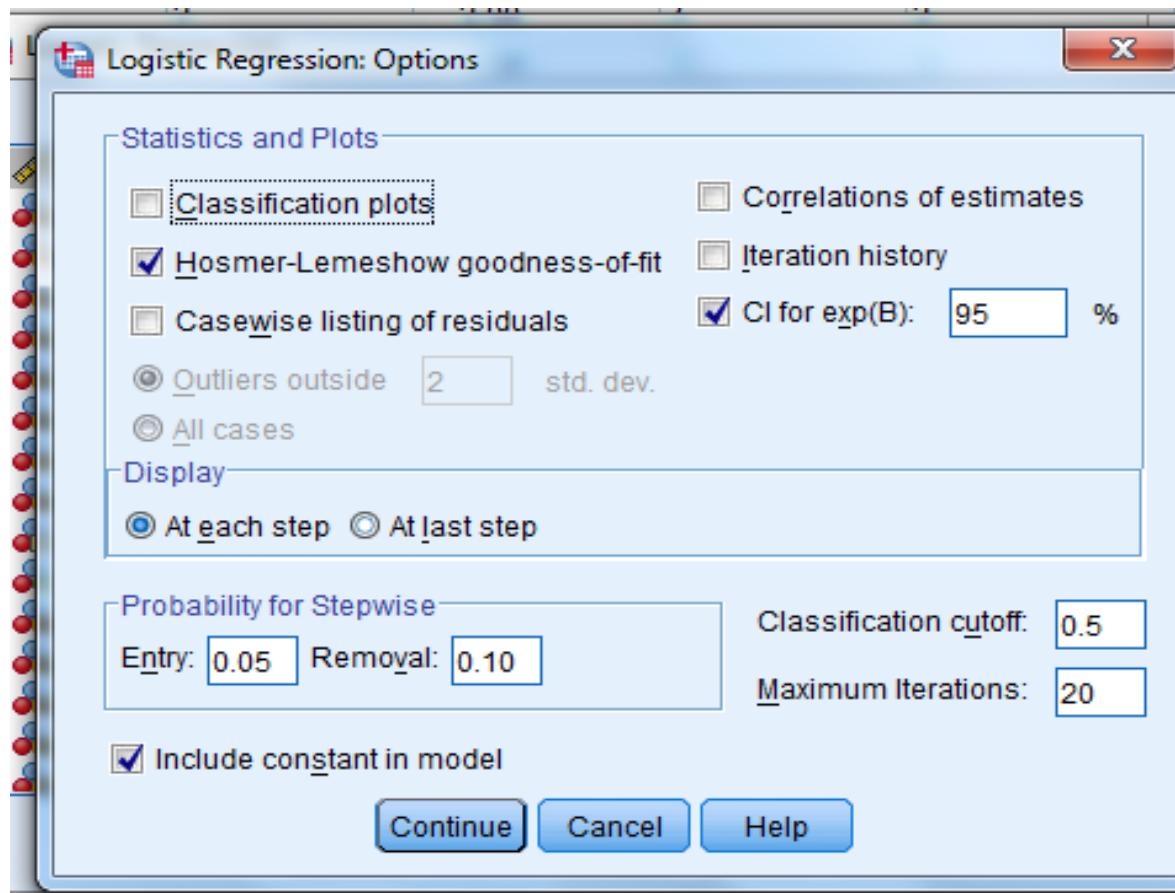
- Reference level should be defined by clicking “**categorical**” and choose either last or first



Tadesse A.

We can change the reference level by choosing either last or first, then click "change" <sup>43</sup>

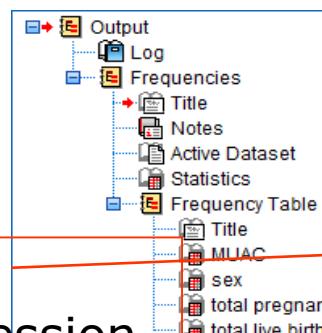
# Binary logistic regression



# Univariate logistic regression

- First run the frequency and check the distribution of each explanatory variable over the dependent variable

This tells us whether the Variable is eligible for regression



Frequency Table

MUAC				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid <11.49	6016	74.7	77.8	
>11.5	1721	21.4	22.2	
Total	7737	96.1	100.0	
Missing System	313	3.9		
Total	8050	100.0		

sex

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid .00	3944	49.0	49.0	49.0
1.00	4106	51.0	51.0	100.0
Total	8050	100.0	100.0	

total pregnancies

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	1804	22.4	22.4	22.4
2-4	3585	44.5	44.6	67.0
>4	2651	32.9	33.0	100.0
Total	8040	99.9	100.0	
Missing System	10	.1		
Total	8050	100.0		

Tadesse A.

# Univariate logistic regression

	Variables in the Equation						P-value	95% C.I. for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper	
Step 1 <sup>a</sup>	.219	.055	15.970	1	.000	1.245	1.118	1.387	
Constant	-1.367	.040	1147.459	1	.000	.255			

a. Variable(s) entered on step 1: sexChild.

- As we see from p-value (<0.05) or
- 95% CI (failed to consist 1), sex of the child has significant association with birth weight of the new born
- The Odds ratio here is **unadjusted**, which may be affected by confounding variables
- However, in the univariate analysis we only focus on 20% p-value

# Multivariable logistic regression

- Now we are going to include all potential explanatory variables in the model simultaneously

Explanatory variables with codes used for the analysis

Categorical Variables Codings

	Frequency	Parameter coding	
		(1)	(2)
total live births	1	1.000	.000
	2-4	.000	1.000
	>4	.000	.000
total pregnancies	1	1.000	.000
	2-4	.000	1.000
	>4	.000	.000
sex	.00	.000	
	1.00	1.000	

Hosmer and Lemeshow Test

→

Step	Chi-square	df	Sig.
1	7.963	8	.437

Goodness of fit test of the model can be checked using Hosmer and Lemeshow test (higher p-value---good model)  
Tadesse A.

# Multivariable logistic regression

- Here, the cutoff value is set at 0.05 p-value

The resulting odds ratio is **adjusted** and less affected by confounding variable

	Variables in the Equation					P-value	Adjusted odds ratio	
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
Step 1 <sup>a</sup>	.224	.055	16.433	1	.000	1.251	1.122	1.393
sexChild(1)								
CatGravida			2.581	2	.275			
CatGravida(1)	-.550	.349	2.485	1	.115	.577	.291	1.143
CatGravida(2)	-.139	.205	.463	1	.496	.870	.582	1.300
parity								
parity(1)	-.078	.348	.050	1	.823	.925	.468	1.829
parity(2)	.035	.207	.029	1	.865	1.036	.691	1.554
momage	.001	.006	.024	1	.876	1.001	.989	1.013
Constant	-1.223	.209	34.152	1	.000	.294		

a. Variable(s) entered on step 1: sexChild, CatGravida, parity, momage.

# Survival Analysis

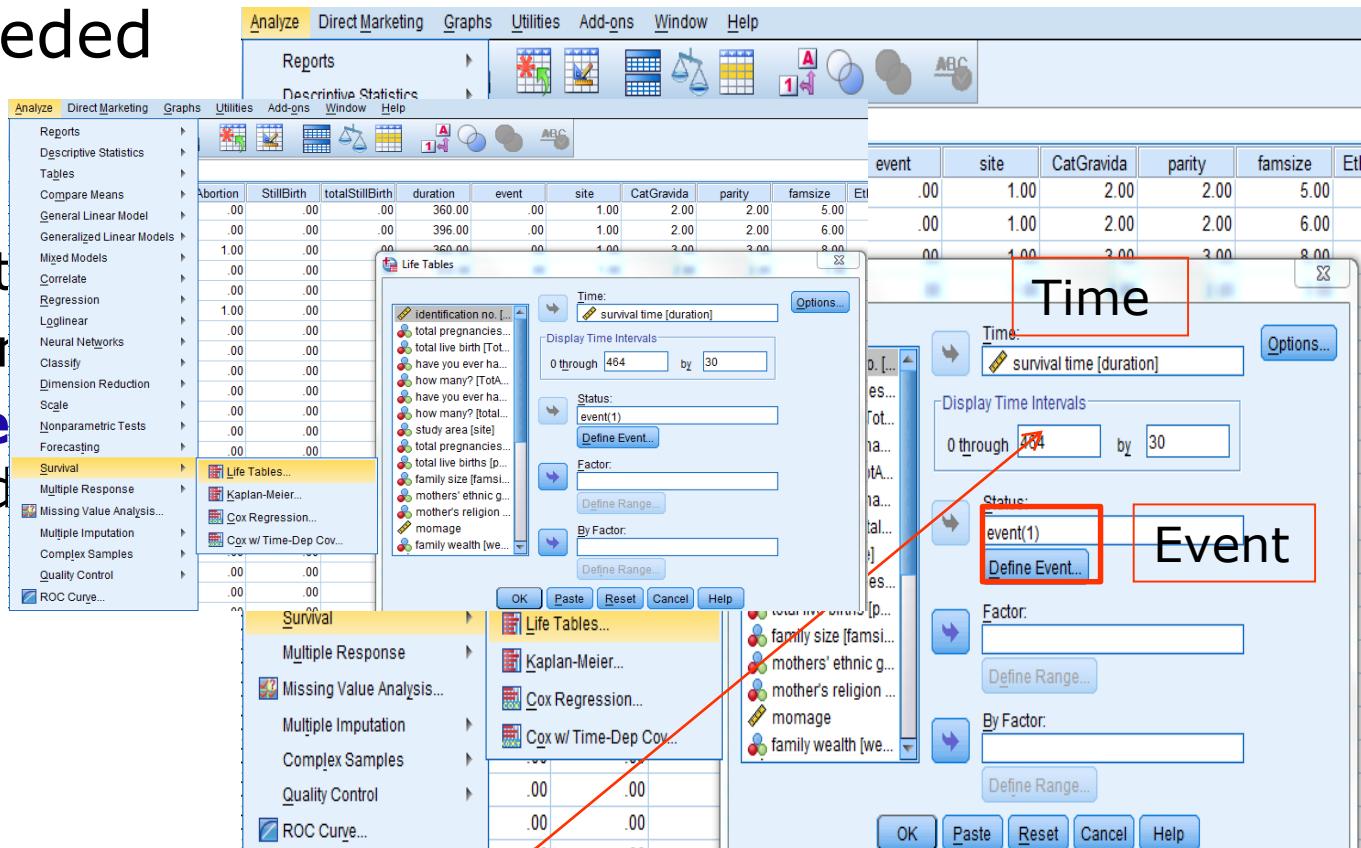
---

- Here the situation is different from what we have seen previously
  - The outcome variable is time-to-event like
  - Time-to-recovery from diseases
  - Time-to-discharge
  - Time-to-diseases progression
  - Time-to-death
  - etc
- All the examples take time and the occurrence of an event
- In this case we use survival analysis

# Life table

Quantities needed are:

- Time
- Max time (interval)
- Event of interest
- Click “**define event**” and click the value for the event
- Click “**OK**”



Max time  
Tadesse A.

# Output

Survival Variable: survival time

Cumulative probability tells us  
the probability of surviving at time t

Interval Start Time	Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval
0	8050	102	7999.000	216	.03	.97	.97
30	7732	6	7729.000	91	.01	.99	.96
60	7635	116	7577.000	82	.01	.99	.95
90	7437	9	7432.500	62	.01	.99	.94
120	7366	96	7318.000	42	.01	.99	.94
150	7228	7	7224.500	48	.01	.99	.93
180	7173	81	7132.500	35	.00	1.00	.93
210	7057	12	7051.000	45	.01	.99	.92
240	7000	72	6964.000	34	.00	1.00	.92
270	6894	17	6885.500	33	.00	1.00	.91
300	6844	150	6769.000	30	.00	1.00	.91
330	6664	912	6208.000	30	.00	1.00	.90
360	5722	5636	2904.000	1	.00	1.00	.90
390	85	77	46.500	0	.00	1.00	.90
420	8	7	4.500	0	.00	1.00	.90
450	1	1	.500	0	.00	1.00	.90

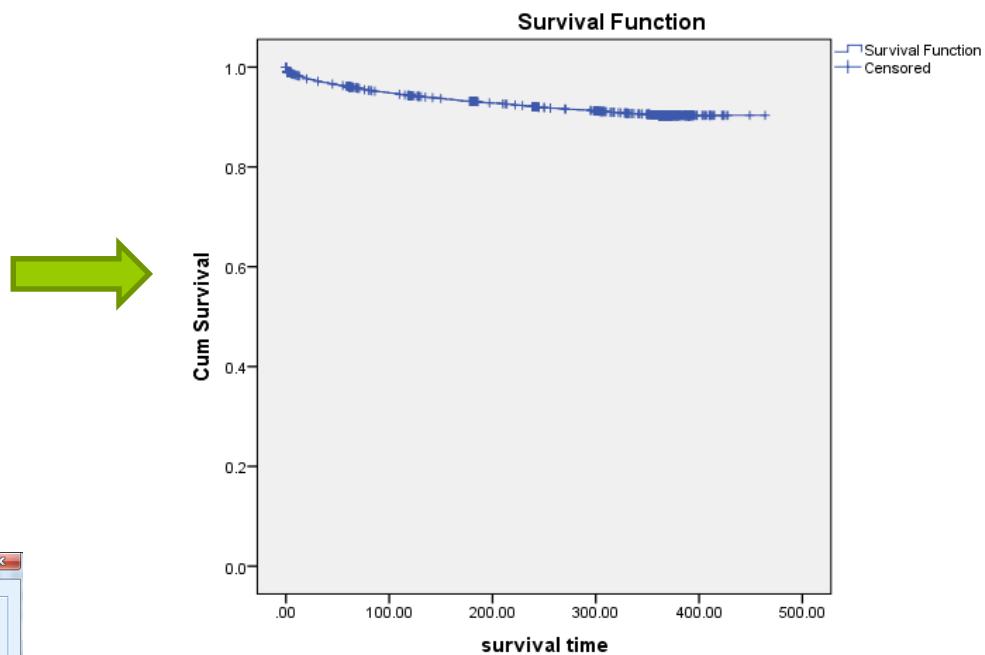
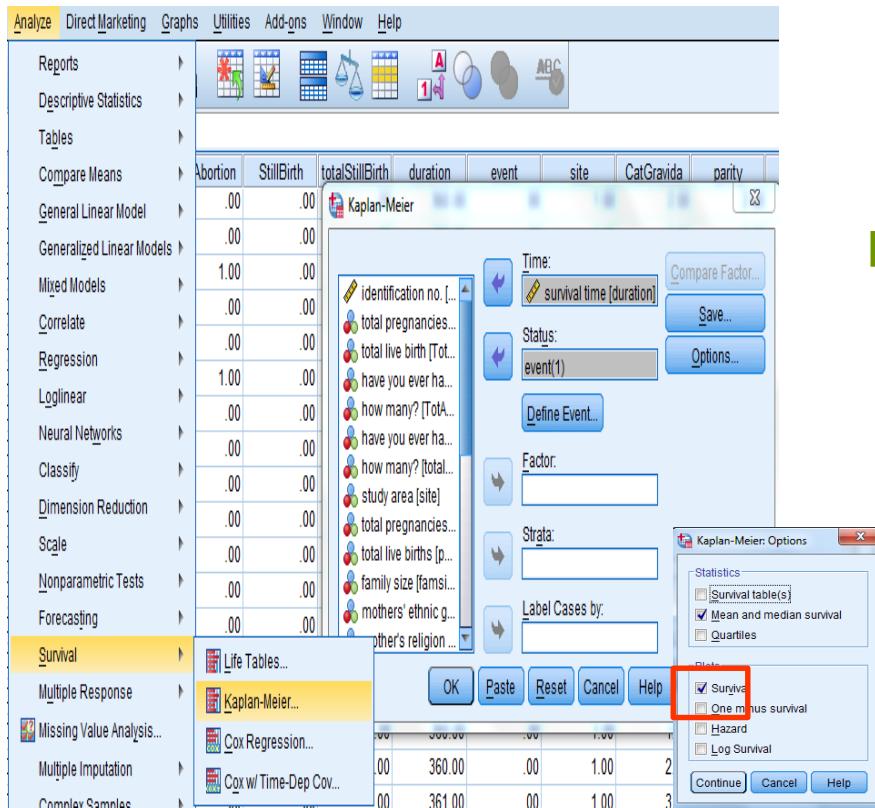
a. The median survival time is 450.0000

Tadesse A.

51

# Kaplan-Meier curve

## □ To construct life table



## Cox-Regression

---

- It helps us to predict the effect of covariates on time to even outcome variable
  -
- This model is semi-parametric
- Hazard ratio, 95% CI and p-value can be computed from the model
- As logistic regression, we use both univariable and multivariable cox-regression to estimate unadjusted and adjusted hazard ratio

# Unadjusted hazard ratio

- To compute the estimates, we need to define
  - Time
  - Event
  - Covariates
- Reference level should be defined by clicking “**categorical**” and choose either last or first

The screenshot shows two overlapping SPSS dialog boxes. The main dialog is titled "Cox Regression" and includes fields for "Time" (survival time [duration]), "Status" (event(1)), and "Categorical..." (highlighted). It also has tabs for "Plots...", "Save...", "Options...", and "Bootstrap...". Below these are sections for "Block 1 of 1" and "Covariates:" (containing "momage" and "sexChild"). The "Method" is set to "Enter". The "Cox Regression: Define Categorical Covariates" dialog is overlaid on the main one. It lists "Covariates:" (momage) and "Categorical Covariates:" (sexChild(Indicator)). It also includes "Change Contrast" options (Contrast: Indicator, Change), and "Reference Category" (Last, First). Both dialogs have "OK", "Paste", "Reset", and "Cancel" buttons at the bottom.

Time: survival time [duration]  
Status: event(1)  
Categorical...  
Plots...  
Save...  
Options...  
Bootstrap...  
Covariates:  
momage  
sexChild  
Method: Enter  
Covariates:  
momage  
Categorical Covariates:  
sexChild(Indicator)  
Change Contrast  
Contrast: Indicator  
Reference Category: Last  
First

# Output

	B	SE	Wald	df	Sig.	Exp(B)	P-value	
							Variables in the Equation	
							95.0% CI for Exp(B)	
sexChild	-.271	.079	11.882	1	.001	.763	.654	.890
weight	-.001	.000	103.298	1	.000	.999	.999	.999
CatGravida			.795	2	.872			
CatGravida(1)	.408	.522	.611	1	.434	1.503	.541	4.179
CatGravida(2)	.296	.351	.711	1	.399	1.345	.676	2.675
parity			.531	2	.767			
parity(1)	-.314	.520	.365	1	.546	.731	.264	2.024
parity(2)	-.250	.353	.502	1	.479	.779	.390	1.555

Adjusted Hazard ratio

Here is the CI and can be used to test the if significant

# Stata application

The screenshot shows the Stata SE 12.0 interface. A red arrow points from the title bar "Stata/SE 12.0 - C:\Users\lucp8319\Desktop\Biostat\_2014\Biostat\_LectureNotes\_2014\_2015\Data\infant\_growth\_data\_baseline.dta - [Results]" down to the "Statistics" menu. The "Binary outcomes" option is selected. A second red arrow points from the "Variables" table on the right towards the "Logistic regression" option in the dropdown menu. The "Variables" table lists various demographic and health variables. A blue text overlay "Possible model options" is positioned over the "Variables" table.

Stata/SE 12.0 - C:\Users\lucp8319\Desktop\Biostat\_2014\Biostat\_LectureNotes\_2014\_2015\Data\infant\_growth\_data\_baseline.dta - [Results]

Statistics User Window Help

Summaries, tables, and tests

Linear models and related

Binary outcomes

Ordinal outcomes

Categorical outcomes

Count outcomes

Exact statistics

Endogenous covariates

Sample-selection models

Multilevel mixed-effects models

Generalized linear models

Nonparametric analysis

Time series

Multivariate time series

State-space models

Longitudinal/panel data

Survival analysis

Epidemiology and related

SEM (structural equation modeling)

Survey data analysis

Multiple imputation

Multivariate analysis

Power and sample size

Resampling

Postestimation

Other

Logistic regression

Logistic regression (reporting odds ratios)

Exact logistic regression

Mixed-effects logistic regression

Panel logistic regression

Probit regression

Probit regression with endogenous covariates

Probit regression with selection

Bivariate probit regression

Seemingly unrelated bivariate probit regression

Panel probit regression

Complementary log-log regression

Panel complementary log-log regression

GLM for the binomial family

Heteroskedastic probit regression

Skewed logit regression

Grouped data

Postestimation

Variables

Variable	Label
idno	identification no.
totpreg	total pregnancies
totlb	total live birth
abortion	have you ever ...
totabortion	how many?
stillbirth	have you ever ...
totalstillbirth	how many?
duration	survival time
event	survival outcome
site	study area
catwbt	total pregnancy weight
parity	total live births
famsize	family size
ethnicmom	mothers' ethnic...
religionmom	mother's religion
momage	
wealth	family wealth
weight	weight
length	length
armcircumfere...	
sexchild	sex
deaths	total child deaths
survtime	survival time
marital	current marital ...
catbwt	birth weight
prgnloss	pregnancy loss
tdh	total child deat...
catagemom	recode of mom...
muac	muac

Possible model options

Command

## Exercises 3

---

- Choose appropriate methods for each of your title of research
- Explain your reason for choosing the model
- Use your own data to:
  - compute descriptive statistics
  - Construct appropriate graph/chart
  - Fit appropriate statistical model
  - Interpret the result

# References

---

- MICHAEL RC and ROBERT HF (2003). Introductory biostatistics: Modern Applications Including Bootstrap
- Martin Bland (200). Introduction to Medical Statistics, An, 3<sup>rd</sup> Edition
- Gerald VB., Lloyd DF., Patrick JH, Thomas L, Walter AS, Samuel SW. Biostatistics a methodology for the health sciences Second edition
- An Introduction to Categorical Data Analysis, Second Edition. By Alan Agresti 2007 JohnWiley and Sons, Inc.
- Brajendra C. Sutradhar (2014): Longitudinal Categorical Data Analysis
- Garrett Fitzmaurice (2014): Handbooks of Modern Statistical Methods:Handbook of Survival Analysis
- <http://courses.csusm.edu/resources/spss/>

# The End



# Session VIII

# Ethics in Research



# Learning Objectives

---

At the end of this session the student will be able to:

- Understand the historical development of ethics in health research
- Understand the major ethical principles of research
- Develop consent forms

# Introduction

---

- Ethics
  - moral principles of right and wrong
  - A set of standards that includes both prohibitions against and prescriptions for specific kinds of behavior in research
    - It is not absolute; may vary by person, by time, by place
    - involves a balance between and within principles and practices
- Ethical choices confront us when we provide health care and conduct of research

# Introduction...

---

- Research involving human subjects is defined as:
  - any biomedical, social science, behavioral, or epidemiological activity that involves
    - systematic collection and analysis of data with the intent to generate knowledge or solve problems
- Such research must respect
  - dignity
  - safety
  - rights of research participants

# Historical Development of Medical Ethical

---

## Tuskegee experiment on syphilis (1932–72)

- Studied the natural history of syphilis in 400 African American males
- Motivated by the concern that syphilis might be less harmful in this group
- men afflicted with syphilis to go untreated for a period of 40 years
- Finally
  - ▣ 28 persons had died from the disease,
  - ▣ 100 persons had died from related diseases



# Historical Development ...

---

## Nazi experiments (1939–45)

- was the first among all evil human experimentation
- seventy medical researches were conducted in Nazi concentration camps
- including:
  - ▣ **High-Altitude, Freezing**
  - Malaria, Nerve



# Historical Development ...

## Thalidomide use (1950s)

- In a post-war era when sleeplessness was prevalent, thalidomide was marketed to sleeping pills
- it was not disclosed to patients that the drug was still in the testing phase of the regulatory process
  - ▣ Later discovered that the drug had **teratogenic** effects
    - causing severe deformities in the fetus

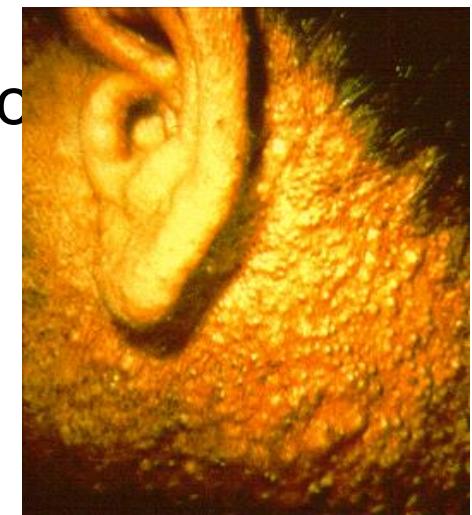


# Historical Development ...

---

## ❑ Cold War Human Radiation Experiment (1944–74)

- Over 4000 secret and classified radiation experiments by
  - ❑ Atomic Energy Commission (AEC)
  - ❑ other government agencies
- believing low dose radiation was not harmful
  - ❑ ignoring
    - ❑ increased risk of cancer
    - ❑ heart disease
    - ❑ neurological disorders
    - ❑ immune system disease
    - ❑ reproductive abnormalities
    - ❑ birth defects, and genetic mutations



## Historical Development ...

---

### □ San Antonio Contraceptive Study (1970)

- involved randomizing subjects between active and placebos contraceptive pills
- The women were not informed as they were the subjects of this type of research
- →There were high numbers of unplanned pregnancies in the placebo group.

# Response to Research Abuses

---

- Since 1945, Various codes have been adopted by different organizations
  - The best known are:
    - i. Nuremberg Code of 1947
    - ii. Universal Declaration of Human Rights (1948)
    - iii. Declaration of Helsinki (1964)
    - iv. National research act (1974)
    - v. Belmont Report (1979)

## i. Nuremberg Doctors Trial (1946)

- led to the formulation of the [Nuremberg Code](#) (1948)
  - 23 persons were charged
  - Of these, 7 were sentenced to death by hanging, 9 were given prison terms, and 7 were found not guilty
- It outlined ethical principles required for research
  - Now regarded as a milestone toward the establishment of a permanent international court



## ii. Universal Declaration of Human Rights (1948)

---

- ❑ The United Nations (UN) came into being in 1945 by 51 countries
- ❑ It is the most universal human rights document in existence
- ❑ Consists of 30 articles elaborated in
  - subsequent international treaties
  - regional human right instruments
  - national constitutions and law
- ❑ Ethiopia becomes a member in 1945

### iii. Declaration of Helsinki (1964)

---

- World Medical Association drafted in 1964 (18<sup>th</sup> Assembly)
  - Has been revised 7 times and the current one is in 2013 (50<sup>th</sup>) - 10 younger UoG
  - It consists of 2 preamble and 35 general principles (37)
    - developed on the Nuremberg Code
- It adds three key points on NCode:
  1. Higher priority for the subject than society
  2. Every subject should get the best known treatment
  3. independent review of all human subject research
- It is the origin of Institutional Review Board (IRB).

## iV. National Research Act (1974)

---

- Response for Tuskegee Syphilis Study
- The Commission had four goals that it needed to analyze
  1. boundaries between biomedical and behavioral research
  2. assessing the risks and benefits of the appropriateness of research
  3. determining appropriate guidelines for how human subjects can be chosen for the participation
  4. defining what informed consent is in each research setting
- Recommendation for Belmont report

## v. Belmont Report

---

- **Ethical Principles and Guidelines** for the protection of Human Subjects of Research
  - **Attempts to summarizes the basic ethical principles** identified by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research
- 
- Consists three ethical principles:
    - Respect for person
    - Beneficence
    - Justice

## Ethics in Health Research

---

- Health research that uses human being as a subject has benefits and risks to the society
- It is beneficial to the society as it ensures good health
- It is risk in that the human participant of research may be exposed to undesirable consequences
- In order to strike a balance between these two conflicting societal interests, countries have developed laws and ethics to promote health research

## Ethics in Health...

---

- Risks in general fall into three categories
  - Invasion of Privacy
  - Breach of Confidentiality
  - Study Procedures
- When assessing risk associated with participation consider
  - The Probability of Harm
  - The Magnitude of Such Harm
- There are three widely recognized principles
  - Respect for Persons
  - Beneficence/Non-Maleficence
  - Justice/Non-Exploitation

# Respect for Persons

---

- **Autonomy:** Each individual:

- Is unique and free
  - Has the right and capacity to decide
  - Has value and dignity
  - Has the right to **informed consent**

- **Protection for vulnerable persons**

- Special protections for those whose decision making capacity is impaired or diminished

- **Application:** Informed Consent

- The consent process must include :
    - Information
    - Comprehension, and
    - Voluntary participation

# Beneficence

---

- **Protection of the study participants** is the most important responsibility of the researcher
  - **Researchers must:**
    - Protect the physical, mental and social well-being
    - Minimizes physical and social risks
    - Maximize the possible benefits
    - Retain the community perspective
- **Application: Assessment of risks and benefits**
  - The nature and scope of risks and benefits must be assessed in a systematic way

# Justice

---

- The principle that calls for **fairness**
- **Research must:** **5 formulations to each person:**
  1. an equal share
  2. according to individual need
  3. according to individual effort
  4. according to societal contribution
  5. according to merit
- **Application: Selection of participants**
  - There must be fair procedures and outcomes in the selection of research participants

# Elements of informed consent

---

1. competence
2. disclosure
3. understanding
4. voluntariness
5. Consent

In other words,

One gives an informed consent if (and perhaps only if) one:

- is competent to act,
- receives a thorough disclosure,
- comprehends the disclosure,
- acts voluntarily, and
- consents to take part in the research

# Consent by surrogates cont...

---

- Parents are considered surrogates for their minor children
- Spouses for one another
- Adult children for parents when parents are lacking
- Adult grand children for grand parents

# Contents of the consent form

---

- The consent form need to include the following
  - Identification of the researcher
  - Identification of the sponsoring institution
  - Indication of how the participants were selected
  - Identification of the purpose of the research

# Contents cont...

---

- Identification of the benefits for participating
- Identification of the level and type of participant involvement
- Notation of risks to the participant
- Guarantee of confidentiality to the participant
- Assurance that the participant can withdraw at any time
- Provision of names of persons to contact if questions arise

# Confidentiality

---

- Confidentiality is necessary in diagnostic, therapeutic and research context
- Confidential information is both private and voluntarily imparted in confidence and trust
- If research subject authorizes release of the information to others, then there is no violation of rights or confidentiality

## Confidentiality cont..

---

- Research may involve collecting and storing data relating to individuals and groups
- such data, if disclosed to third parties, may cause harm or distress.
- investigators should make arrangements for protecting the confidentiality of such data by:
  - omitting information that might lead to identification of individual subjects
  - limiting access to the data, or by other means.
- When personal identifiers remain on records used for a study, investigators should explain why this is necessary and how confidentiality will be protected.

# Confidentiality cont...

---

## Unlinked information

- **Refers to information** that can not be linked, associated or connected with the person to whom it refers.
- As this person is not known to the investigator, confidentiality is not at stake and the question of consent does not arise.

# Confidentiality cont...

---

**Linked information** may be:

- **anonymous** – when the information can not be linked to the person to whom it refers except by a code or other means known only to that person, and the investigator can not know the identity of that person
- **non-nominal** –when the information can be linked to the person by a code which is not a personal identifier and which is known to the person and the investigator

# Dissemination and Utilization of Results

---

- **Briefly describe the dissemination plan**
  - Feedback to the community
  - Feedback to local authorities
  - Identify relevant agencies that need to be informed
  - Scientific publication
  - Presentation in meetings/conferences
- **Briefly describe how the study results can be best translated into application**

# Work plan

---

- Work plan summarizes (in a table, chart, graph) the various components of a research project and how they fit together.
- Includes:
  - Tasks to be performed
  - When the task will be performed
  - Who will perform the task (identify human resource needed for each task)
  - Number of staff needed to perform the task

# Work plan cont..

---

## A work plan can serve as:

- a tool in planning the details of the project activities and later the project funds.
- a visual outline or illustration of the sequence of the project operations.
- a management tool for the principal investigator and/or members of his/her team, showing what tasks and activities are planned, their timing, and when various members will be involved in the tasks;
- a tool for monitoring and evaluation, when the current status of the project is compared to what was foreseen in the work plan.

# **Work plan cont..**

---

## **Ways of presenting a work plan**

Work schedule

GANNT chart

# Work plan cont..

---

## The Work Schedule

- Is a table
- Summarizes:
  - tasks to be performed
  - duration of each activity, and
  - staff responsible.
- *The work schedule does not show how various tasks are related, nor give a visual picture of the time schedule.*

# Work plan cont..

---

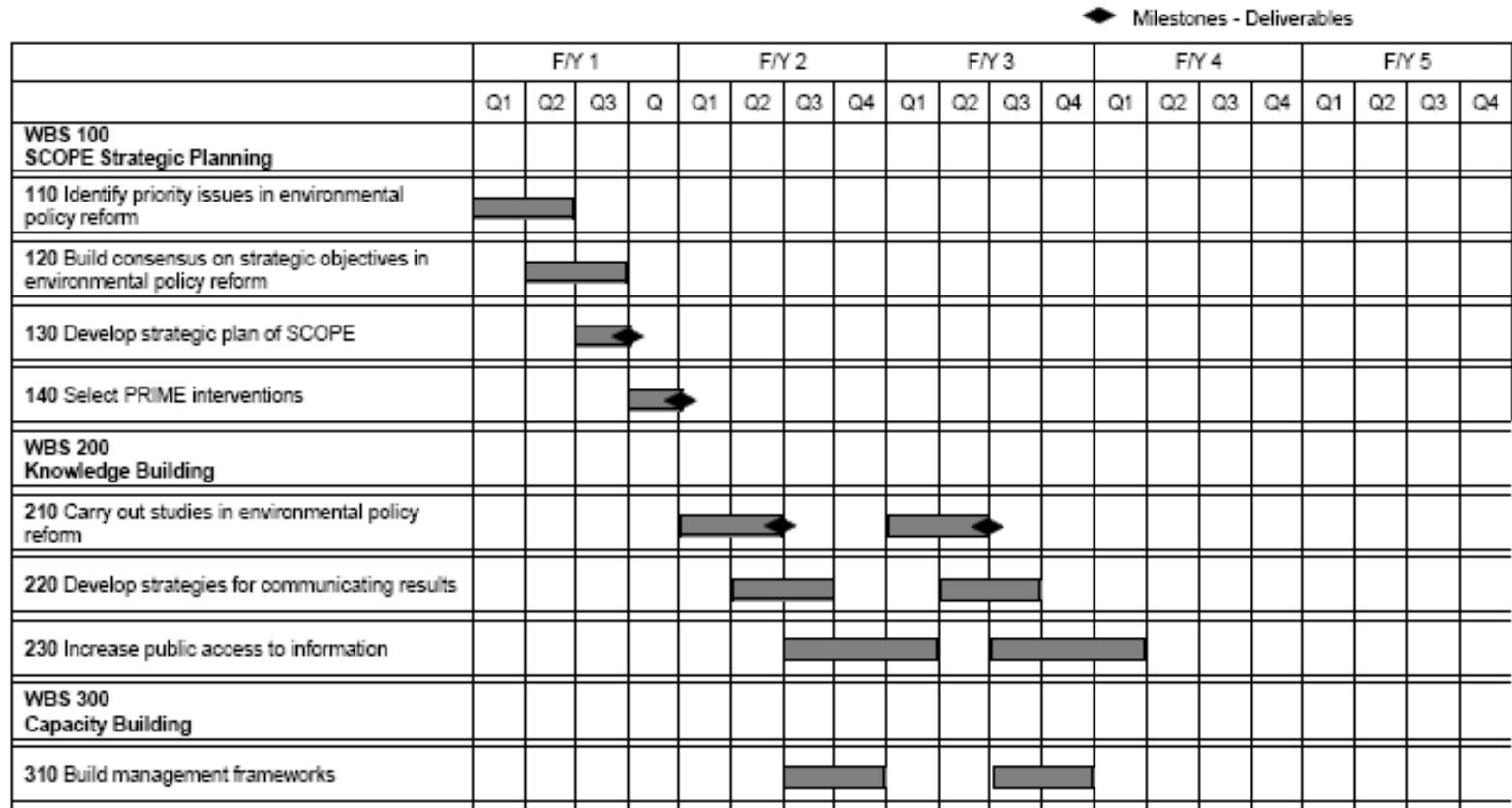
## The GANTT Chart

- Depicts graphically the order in which various tasks must be completed and their duration of activity.

A typical Gantt chart includes the following information:

- The tasks to be performed
  - Who is responsible for each task; and
  - The time each task is expected to take.
  - The length of each task is shown by a bar that extends over the number of days, weeks or months the task is expected to take.
- 
- ***The Gantt chart doesn't show how various tasks are***

# GANNT CHART



# Budget

---

## How should a budget be prepared?

- It is necessary to use the work plan as a starting point.
- Specify, for each activity in the work plan, what resources are required.
  
- Determine for each resource needed the **unit cost** and the **total cost**.
- The budget for the fieldwork component of the work plan will include funds for personnel, transport and supplies.

# Budget cont...

---

## The Budget Format and Justification

- The type of budget format to be used may vary
- Most donor organizations have their own special project forms, which include a budget format.

# Budget cont...

---

## **Advice on budget preparation**

- Include a 5%-10% contingency fund
- If inclusion of a contingency fund is not allowed, an alternative is to slightly over-budget in major categories.
- " Do not box yourself in too tightly with very detailed categories and amounts, especially if regulations do not allow adjustments afterwards.
- Ask the supervising agency to agree that, if necessary there may be some transfer between 'line items' in the budget.

# Budget cont...

---

## Budget justification

- It is not sufficient to present a budget without explanation.
- The budget justification follows the budget as an explanatory note justifying briefly why the various items in the budget are required.
- Make sure you give clear explanations concerning why items that may seem questionable or that are particularly costly are needed and discuss how complicated expenses have been calculated.
- If a strong budget justification has been prepared, it is less likely that essential items will be cut during proposal review.

# References

---

## Methods of citations in preparing literature review:

### A) Vancouver system

This system have been adopted as standard by over 300 biomedical journals

- **For an article the following information should be noted:**

Author(s)' Surname followed by initials. Title of article.

*Name of Journal*. Year, Volume(number): page numbers of article.

- **Example:** Megabiaw B. Emerging- and re-emerging infections: The societal variables. *International Journal of Infectious Disease*. 2004, 1(2):59-62.

# References cont..

---

## B) The *Harvard System*

- In other journals and books it is common to put the year, between brackets, straight after the name of the author(s).
- This is called the Harvard system.
- Name of the author(s) (year). Title. Place of Publication: Publisher
- Ex- Abram H (1990), 4<sup>th</sup> ed. *Survey methods in community medicine*. Edinburgh: Churchill Livingstone.
- There are more systems in use for referencing to literature.
- Always carefully look what system is used in the journal you are submitting an article to and follow it systematically.

## References cont...

---

- At present many journals use as few punctuation marks as possible.
- Examples: In Harvard style, this looks as follows:
- Abramson JH (1990) 4th ed. *Survey methods in community medicine*. Edinburgh: Churchill Livingstone.

# References cont...

---

- When you use the Vancouver system, you will use consecutive numbers in the text to indicate your references
- At the end you will then list your references in that order, using the format described above
- In Harvard System, put the surname of the author, year of publication and number(s) of page(s) referred to between brackets, (E.g. Shiva 1998:15-17)
- If this system of citation is used, the references at the end of the proposal, should be listed in Alphabetical order.

# Annex

---

- Include in the appendices of your proposal any additional information you think might be helpful to a proposal reviewer.
- For example, include:
  - Biographical data on the principal investigator
  - The study questionnaire if you have it.
  - The consent form.
  - A copy of the approval from the Institutional Review Board.

# Assignment

---

- Write

- Ethical consideration
- Dissemination of result
- Work plan
- Budget
- Reference

- Section for your proposed research!

# Session IX

# Research proposal writing



# Learning Objectives

---

- At the end of this session the student will be able to:
  - Know all the subsection of the proposal
  - Develop research proposal

# Components of research proposal

---

- ❑ Title
- ❑ Acknowledgment
- ❑ Table of content
- ❑ Acronym
- ❑ Summary
- ❑ Introduction
  - Background
  - Statement of the problem
  - literature review
  - Justification
- ❑ Objective
- ❑ Methodology
- ❑ Ethical considerations
- ❑ Dissemination and Utilization of Results
- ❑ Work plan
- ❑ Cost of the Project
- ❑ References
- ❑ Annex
- ❑ Assurance of the investigator
- ❑ Consent sheet
- ❑ Questionnaire
- ❑ Advisor (approval)

# Title

---

- Should be in line with your general objective
- Should tell readers what your study is about and where it will be done
- *Make sure that it is specific enough to tell the reader what your study is about and where it will be conducted*
- *Should answer the question: what, where, on whom*

NOT: 'A study on community home-based care'

BUT: 'A study on cost and quality of community home-based care for HIV/AIDS patients and their communities in Ethiopia'

You might also consider fancier titles:

'Do We Care? A study on cost and quality of CHBC for HIV/AIDS patients in Ethiopia'

# Summary

---

Should reflect:

- ❑ Problem Statement
- ❑ Research objectives
- ❑ Research design
- ❑ Duration
- ❑ Total Budget
- ❑ *Keep to about 250-300 words*

# Introduction

---

- Can be divided into 3 sections or the concepts of the 3 sections can be merged together
- Sections:
  - Statement of the problem
  - Literature review
  - Justification of the study

## Includes:

- Brief description of the socio-economic & cultural characteristics
- Overview of health status and health care system
- The **nature** of the problem; and the discrepancy

## What are the possible sources of information?

---

- Individuals, groups, and organizations
- Published information (books, articles, indexes, abstract journals); and
- Unpublished information (other research proposals in related fields, reports, records, computer data bases)
- Computerized search: Hinari, google (title), Medline, pubmed ([www.pubmed.gov](http://www.pubmed.gov)), etc.

# How to write a review of literature?

---

- Take your statement of problem as a framework.
- Organize your notes in groups of related statements according to which aspect of the problem they touch upon, e.g., community factors, service factors.
- Decide in which order you want to discuss the various issues.
- If you discover you have not yet found literature, make a special effort to find this literature.
- If there is no literature, this supports your justification for conducting the study
- All facts you mention need a source/reference/except some general and well-known statements

# Justification

---

- Why your research is needed?
- E.g
  - Others not large enough
  - Different populations ????
  - Different intervention
  - Different tools, ....

# Objectives

---

- ❑ A research objective summarize what is to be achieved by the study.
- ❑ Identify **General and Specific objectives.**

# Methodology

---

## Possible sub-sections of the methodology:

- the study design
- Setting
- study population
- variables
- sample size and sampling strategies
- Data collection: what type, how (tool/procedure), who (enquires/responds), where (geography/enquiry place), when (period/time)
- Data quality control issues
- Data management: coding, entering, cleaning, storing, recoding (software to be used)
- Data analysis: choice of statistical methods
- Ethical Concerns

# Study design

---

## □ Design options:

- Quantitative
- Qualitative
- mixed

# Setting

---

- Health institutions
- Schools
- community

# Study Population

---

- ❑ Who are your study subjects?
- ❑ Inclusion and exclusion criteria

# Sample size and sampling strategies

---

- Describe how the sample size is determined
- Describe the methods of sample selection
- If needed, use diagrams to simplify the sample selection process

# Variables

---

- Dependent/independent variables
- Operationalizing variables (making them measurable)

# Operational definitions

---

- Many variables can easily be measured
- For some variables it is sometimes not possible to find meaningful categories unless the variables are made operational with one or more indicators
- Operationalizing variables means that you make them measurable

# Data collection

---

- ❑ Data collection techniques and tools
- ❑ Who will collect the data?
- ❑ Who will supervise the data collection?
- ❑ How long will take the data collection? etc

# Data Quality Control Issues

---

Describe/provide:

- Operational definitions of crucial concepts
- Selection and training of field workers/research staff
- Field testing the research methods and tools
- Supervision and quality control

# Data management

---

- Pretest
- Pilot study
- Data management
  - **Data entry**
  - **Data coding**
  - **Data Cleaning**
  - Data analysis

# Ethical Considerations

---

- Professional obligation to safeguard the safety of study subjects
- Refer to national and international guidelines
- Describe potential ethical concerns and mechanisms to minimize harm and maximize benefits

***Every research can potentially cause ethical concerns!!***

# Work plan

---

## **Ways of presenting a work plan**

Work schedule

GANNT chart

PERT chart.

# Budget

---

## How should a budget be prepared?

- It is necessary to use the work plan as a starting point.
- Specify, for each activity in the work plan, what resources are required.
- Determine for each resource needed the **unit cost** and the **total cost**.
- The budget for the fieldwork component of the work plan will include funds for personnel, transport and supplies.

# References

---

- A) Vancouver system
- B) The *Harvard System*

# Annex

---

- ❑ Include in the appendices of your proposal any additional information you think might be helpful to a proposal reviewer.
- ❑ For example, include:
  - Biographical data on the principal investigator
  - The study questionnaire if you have it.
  - The consent form.
  - A copy of the approval from the Institutional Review Board.

# Ethics review ...

Institute of Public Health  
College of Medicine and Health Sciences  
University of Gondar

## **ETHICAL REVIEW FORMAT**

Name of Students \_\_\_\_\_

Title of the proposed study clearly stated \_\_\_\_\_  
\_\_\_\_\_

S.no	Content	Comment
1	Are the objectives of the study clearly stated?	
2	Are the methods scientifically sound? Are the methods appropriate to meet the objectives?	
3	Are ethical issues well addressed?	
3.1	Is consent form and participant information sheet attached?	
4	For experimental or intervention studies, does the consent form contain all the necessary information?	
5	(For experimental or intervention) are safety procedure in the use of vaccine, drugs and other biological products well described?	
6	Other comments	

### Recommendation

1. Accept without revision
2. Accept with minor revision
3. Accept with major revision
4. Reject the proposal

Reviewer's name \_\_\_\_\_

Signature \_\_\_\_\_

Date: \_\_\_\_\_

# Project

---

- Develop full proposal for your title of research?
  
- **Submission date: One day before the exam**

# Session X

# Paper writing



Tadesse Awoke (PhD)

# Sections of scientific Paper

---

- Title
- Abstract
- Introduction
- Methods
- Results
- Discussion
- Conclusion
- Reference
- Acknowledgement

# Cover page

---

Should contain:

- the title
- name of the authors with their titles and positions
- the institution that is publishing the report
- the month and year of publication

# Title

---

Purpose: To provide a brief , informative summary that will attract your target audience

The screenshot shows a digital journal article interface. On the left, there's a sidebar with icons for messaging, bookmarks, and a PDF download. The main content area has a header with the journal name 'Archives of Public Health' and a DOI link. Below the header, there are two tabs: 'RESEARCH' (which is highlighted) and 'Open Access'. The main title of the article is 'Undernutrition and associated factors among adults living with Human Immune Deficiency Virus in Dembia District, northwest Ethiopia: an institution based cross-sectional study'. The authors listed are Anbesaw Mitiku<sup>1</sup>, Tadesse Awoke Ayele<sup>2</sup>, Mekonen Assefa<sup>3</sup> and Amare Tariku<sup>4\*</sup>. There's also a 'CrossMark' logo. At the bottom, there are sections for 'Abstract' and 'Background'.

Mitiku et al. *Archives of Public Health* (2016) 74:33  
DOI 10.1186/s13690-016-0143-y

RESEARCH Open Access

CrossMark

**Undernutrition and associated factors among adults living with Human Immune Deficiency Virus in Dembia District, northwest Ethiopia: an institution based cross-sectional study**

Anbesaw Mitiku<sup>1</sup>, Tadesse Awoke Ayele<sup>2</sup>, Mekonen Assefa<sup>3</sup> and Amare Tariku<sup>4\*</sup>

**Abstract**

**Background:** Appropriate dietary intake determines the disease progression and success of Anti-Retroviral Therapy

# Abstract

---

- The abstract will be the first (and for busy decision makers most likely the only) part of your study that will be read
  
- Writing abstracts demands thorough reflection and is time consuming
- **Purpose** : To highlight key points from the major sections of the abstract

# Components of the abstract

Component of abstract	Abstracted from
Brief description of the problem (Why this study was needed)	Introduction
Main objectives of the study	Introduction
Place of the study	Methods
Type of the study & methods used	Methods
Main findings	Results
Conclusions and	Discussion

# Abstract cont...

---

- Abstract can be structured or unstructured (always check the instruction of the journal)
- Remember that only few words are allowed to be included in the abstract (usually 250-300 words)
- Emphasize on what is new and useful

# Abstract

The screenshot shows a PDF document open in a viewer. The top menu bar includes File, Edit, View, Document, Tools, Window, and Help. The toolbar below has icons for file operations like Open, Save, Print, and a magnifying glass for search. A status bar at the bottom indicates '1 / 8' pages and '121%' zoom level. A message in the center says 'You are viewing this document in PDF/A mode.' On the left, a sidebar titled 'Bookmarks' lists the document structure:

- Abstract
  - Background
  - Methods
  - Results
  - Conclusions
- Background
- Methods
  - Study area and period
  - Sample size and sampling procedure
  - Data collection instruments and procedures
  - Data processing and analysis
- Results
  - Socio-demographic and economic characteristics
  - Medical related characteristics
  - Prevalence of undernutrition and dietary pattern related characteristics

The main content area is titled 'Abstract'. It contains the following text:

**Background:** Appropriate dietary intake determines the disease progression and success of Anti-Retroviral Therapy (ART). Undernutrition unacceptably increases the risk of mortality among adults living with Human Immune Deficiency Virus (HIV). However in resource limited settings including Ethiopia, many of HIV positive clients lack access to sufficient quantities of nutritious food. There is limited evidences showing the magnitude of undernutrition in this segment of the community, particularly in the rural residents. Therefore, this study aimed to assess undernutrition and associated factors among HIV positive adults attending ART clinic in Dembia District.

**Methods:** An institution based cross-sectional study was conducted in Dembia District from October 1 to 30, 2015. Systematic random sampling technique was used to recruit the study subjects. The anthropometric measurement, Body Mass Index, was computed to determine the nutritional status of the study participants. In order to identify factors associated with undernutrition a multivariable logistic regression analysis was employed. The Adjusted Odds Ratio (AOR) with 95 % Confidence Interval (CI) was calculated to show the strength of association. In multivariable analysis, variables with a P-value of <0.05 were considered as statistically significant.

**Results:** Of the study participants, about 23.2 % [95 % CI: 19.2, 27.2 %] were undernourished in Dembia District. The result of adjusted analysis revealed that, the odds of undernutrition was higher among adults whose age ranged between 18-29 years [AOR = 2.50, 95 % CI: 1.10, 5.69], who had a Cluster of Differentiation (CD4) count less than 200 cells/mm<sup>3</sup> [AOR = 6.21, 95 % CI: 2.97, 12.98], were widowed [AOR = 2.18, 95 % CI: 1.08, 4.40], and anemic [AOR = 3.17, 95 % CI: 1.70, 5.92].

**Conclusions:** The prevalence of undernutrition among HIV positive adults was higher in the study area. Furthermore, being in the age range of 18-29 years, widowed, anemic, and having a CD4 count of less than 200 cells/mm<sup>3</sup> were positively associated with undernutrition. Therefore, efforts should be strengthened to mitigate the higher burden of undernutrition by considering the identified determinants.

**Keywords:** Undernutrition, HIV positive adults, Northwest Ethiopia

# Introduction

---

- Purpose → to explain why your study is needed
- Often requires just three paragraphs
- Objectives are usually the last part of the introduction

# Components of introduction

---

Component	Purpose
Problem statement	What do we need to learn
Literature review	What do we already know
Purpose statement	What will we learn from the study

# Methods

---

- ❑ Purpose → to describe how you collected, organized and analyzed data that are relevant to the purpose of the study
- ❑ Organize into logical subsections that illustrate the steps you took to collect, organize and analyze the data
- ❑ Describe what you did, not what you found
- ❑ Describe original methods in detail; otherwise give references

# Possible subsection headings of the methods

---

- Study design
- Study area
- Study population
- Sample size and sampling method(s)
- Variables of the study
- Data collection procedures
- Data analysis
- Ethical considerations

# Results

---

- ❑ Purpose → to describe the results of data analysis that are relevant to the study
- ❑ Systematic presentation of your findings in relation to the research objective is crucial

# Presentation of findings

---

Description of findings should offer a good combination or triangulation of data from qualitative & quantitative components of the study

# Two different ways of presenting findings

---

- 1) an integrated presentation of all the data by objective
- 2) presentation of data by research instrument (separate analysis is indicated for objectives that are covered by distinct study popn using specific instrument)

# Sequence of presentation of findings

---

- The 1<sup>st</sup> section is usually description of the study population
- Provide information on the problem you studied (size, distribution, characteristics)
- In analytic study, the degree to which different independent variables influence the problem

# Use of tables & figures

---

- Tables & figures need numbers & clear title
- Include only those tables & figures that present main findings & need more elaborate discussion in the text
- Use tables to highlight individual values
- Use figures to highlight trends/relationships
- Check the numbers; provide consistent row & column summations
- In tables, keep lines to a minimum; avoid vertical lines
- Use footnotes to clarify points of potential ambiguity

# Text

---

- Describe what you found, not what you did
- Present results in a logical sequence
- Consider sub-sections
- Less is often better ( a table or figure is worth thousand words)
- Make sure all numbers in text are consistent with tables/figures

# Discussion

---

Purpose → to interpret your results & justify  
your interpretation

# Guidelines for constructing the discussion

---

- Findings can be discussed by objective or cluster of related variables or themes, which should lead to conclusion & possible recommendation
- Focus on the main results
- Emphasize on what is new or different
- Restate the main finding
- Interpret the result in context of literature
- State conclusions & recommendations

# Conclusions & recommendations

---

- ❑ Conclusions should be short, as they have been elaborately discussed in the discussion section
- ❑ Recommendations are usually placed in roughly the same sequence as the conclusion

# Conclusions & recommendations cont...

---

- ❑ Recommendations may at the same time be summarized according to the groups towards which they are directed

## Example

- Policy makers
- Managers at district or lower level
- Staff who could implement the activities
- The community at large

# References

---

- Check the style of reference writing which is recommended by the journal where you want to submit your manuscript
  
- In our university Vancouver system is used

# Annex

---

- ❑ Findings which are less important to be presented in the main section of the paper will be presented here

# Assignment 10

---

- Individual assignment
- For each published paper given to you;
  - I. Criticize all the method section
  - II. Give your comments on the result, discussion, conclusion, recommendation and reference sections

**END!!!**



# Session VI

## Introduction to EPI-info 7



Tadesse Awoke (PhD)

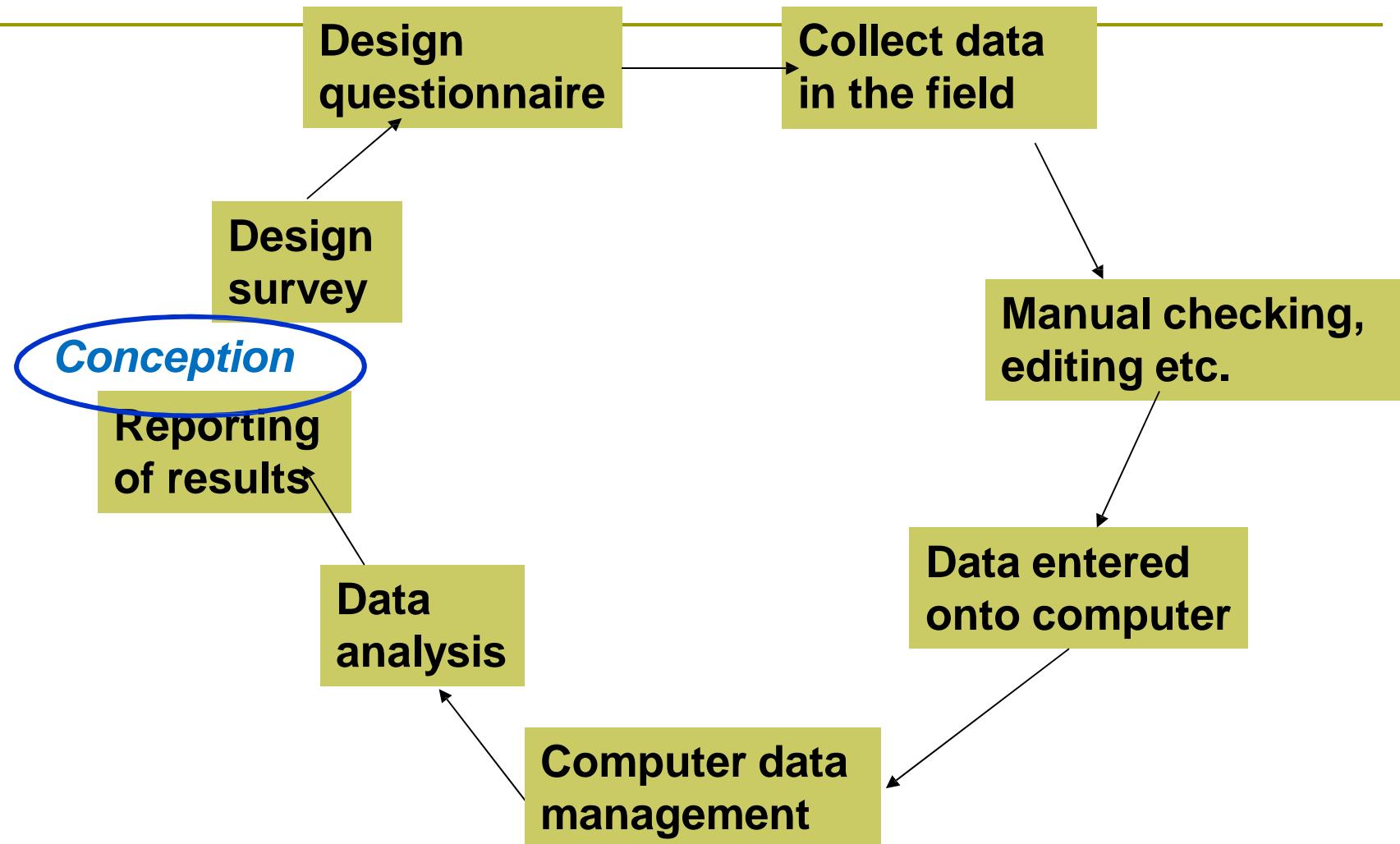
Aug 2018

# Outline

---

- Introduction
- Data management cycle
- Introduction to Epi-Info 7
- Creating forms
- Data entry
- Data analysis

# Data management cycle (1)



# What is Epi Info?

---

- A **free** public domain software package developed by the Centers for Disease Control and Prevention for the global community of medical and public health professionals.
- It can be used to **rapidly**:
  - develop an electronic data entry form
  - enter data into this form
  - analyze data entered into this form or imported from other software packages

## Past, Present, and Future

---

- First version: Epi Info 1 (MS-DOS)
  - Released 1985
- Latest version: Epi Info 3.5.1
  - Released 2008
- Earlier this year, Epi Info became an open source software.
- The new version (Epi Info 7) is currently developed at the CDC's National Center for Public Health Informatics.

# Introduction to Epi-Info 7

---

- Epi Info 7 is a series of freely-distributable tools and utilities for Microsoft Windows for use by public health professionals to conduct
  - outbreak investigations
  - manage databases for public health surveillance and other tasks
  - general database and statistics applications

# Installation

---

- Step 1: Download (get it by CD)
- Step 2: Open the folder and click setup
- Step 3: put the short cut on your desktop
- Step 4: now the EPI-Info 7 started working

# Introduction (1)

---

- Important to:

- Create forms, Data entry, Data analysis and Documentation
- Free program
- Based on past Epi Info, and much developed
- Windows format, can work in Xps and Vista and Window 7 & 8
- No limit on No of observation (tested with >100.000)

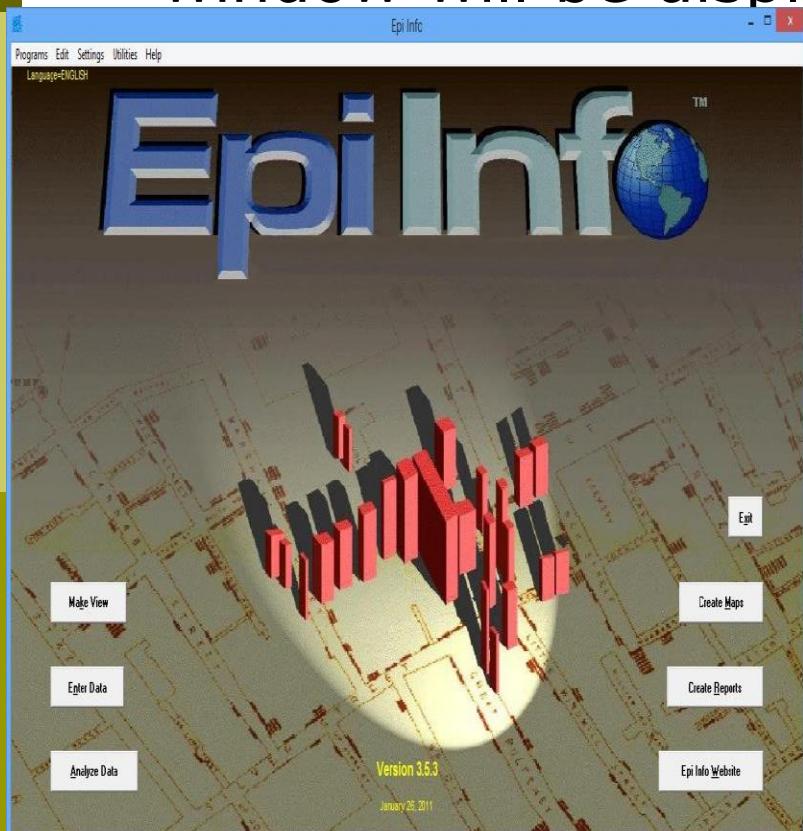
## Introduction(2)

---

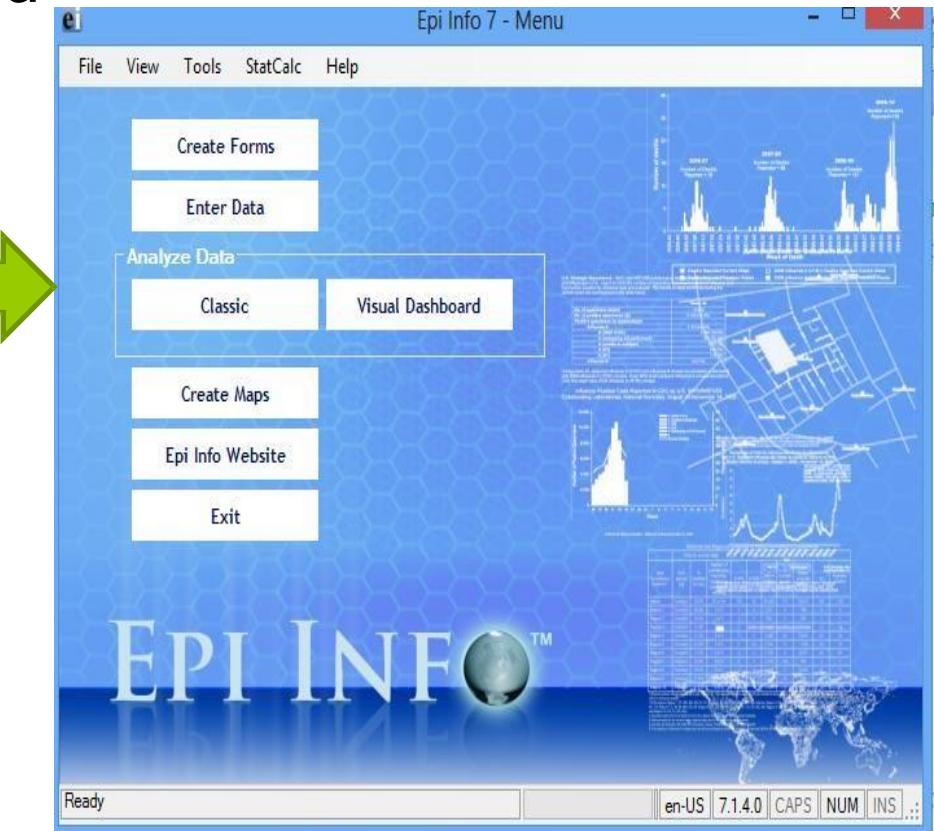
- It enables physicians, epidemiologists, and other public health and medical officials to
  - rapidly develop a questionnaire or form
  - customize the data entry process, and enter
  - analyze data
- Epi Info 7 is free of charge and can be downloaded from the CDC website at <http://www.cdc.gov/epiinfo>

# Introduction (3)

- When you click the short cut, the following window will be displayed



EPI-Info 2002



EPI-Info 7

# Epi-Info 7

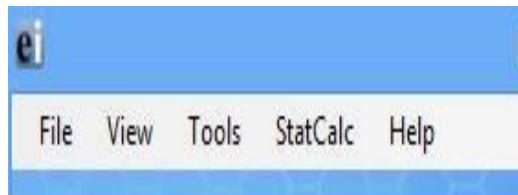
---

## ❑ Application includes

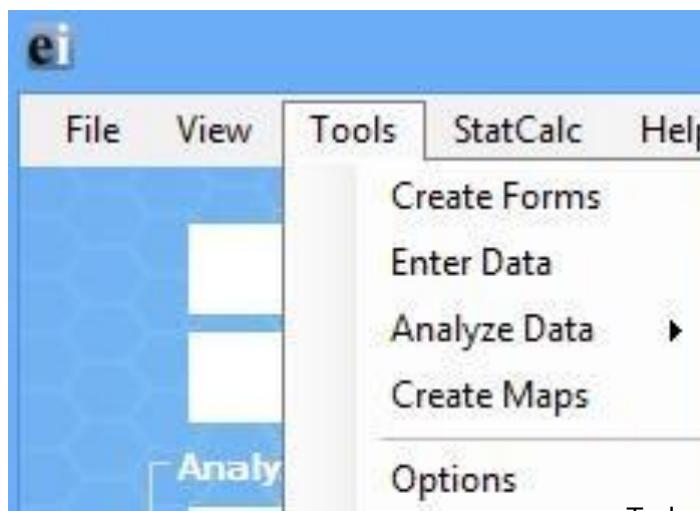
- Creating questionnaire
- Correction of questionnaires, records
- Controlled data entry
- Analysis data (classic and visual dashboard)
- Importing and exporting data
- Creating maps
- Linking to important websites within CDC
- Having Open epi
- etc

# Parts of the software

## □ The pull-down menu



1. Tools can be used to get access of create forms, enter data, analyze data, create maps and options

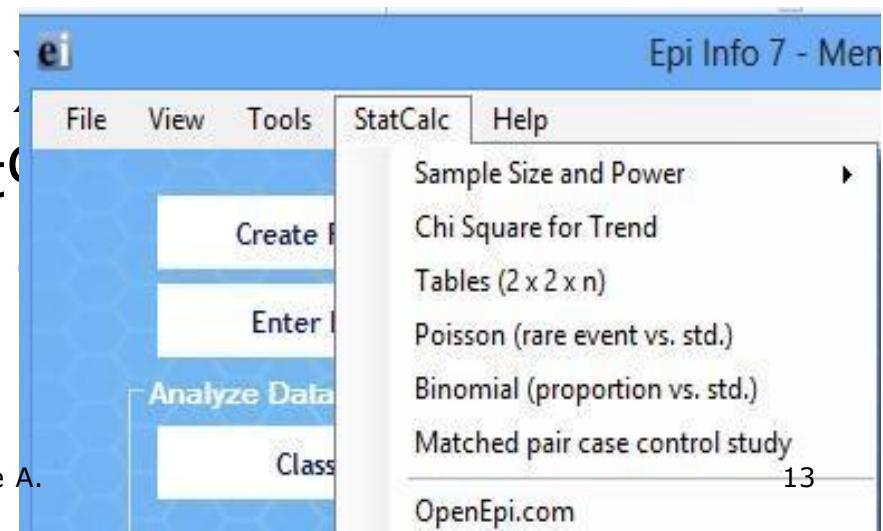


Tadesse A.

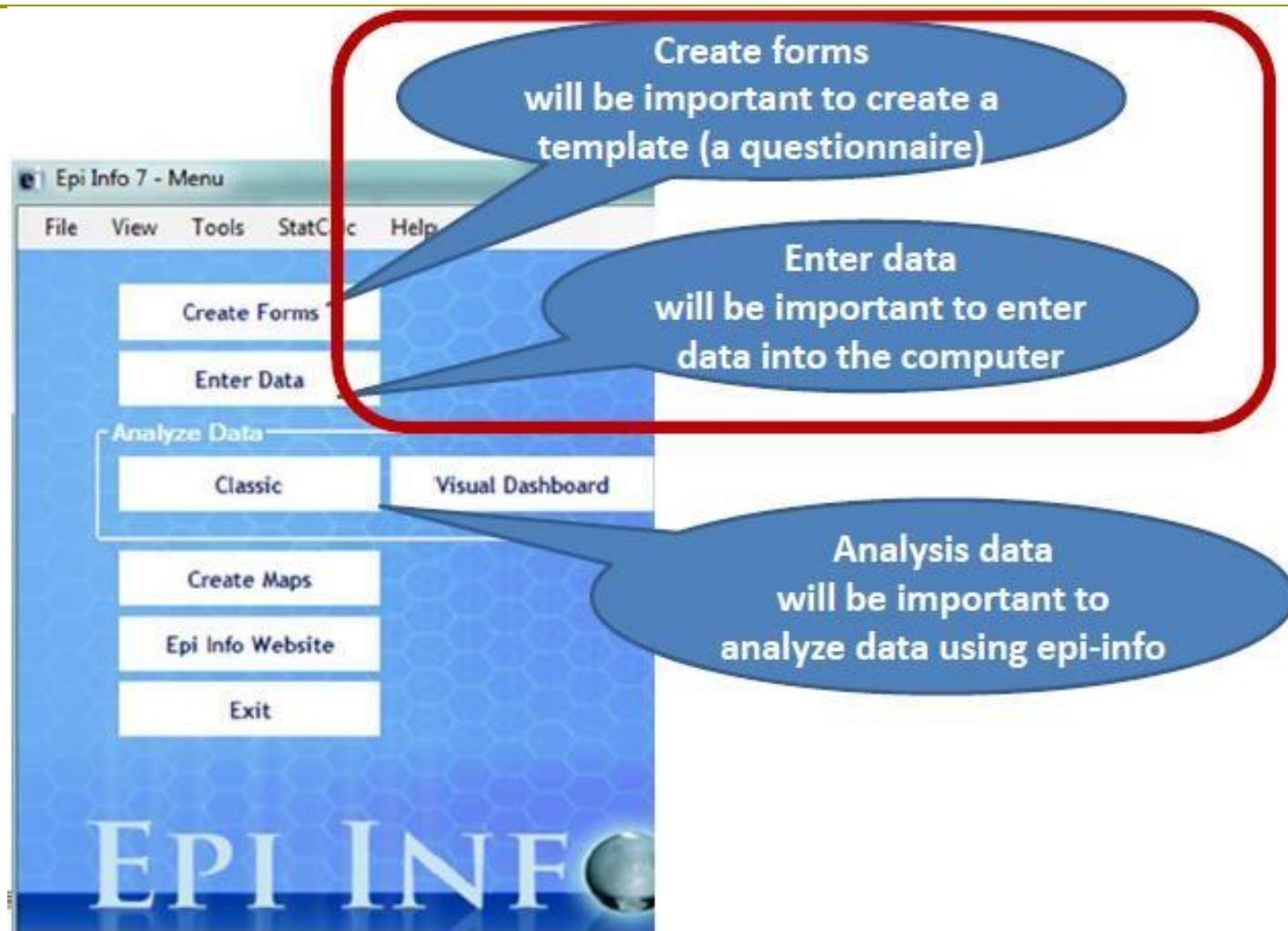
12

## 2. Statcalc

- Which contains:
  - Sample size and power, for
    - Population survey
    - Cohort or analytic cross-sectional design
    - Unmatched case control
- Chi square for trend
- Tables (2 by 2, 2 by n)
- Poisson (rare event vs std.)
- Binomial (proportion vs std.)
- Matched pair case control



# Main menu application



# Creating forms

---

- It is a program necessary to form a template of a questionnaire
- A questionnaire is number of questions (items) each study subject would respond to
- As a template, it will have a space where respondents answered would be entered into
- For each question within the questionnaire, the question,
  - the Field name (variable),
  - Text describing the field and
  - Input definition (number/ letters/ date)

# How to start

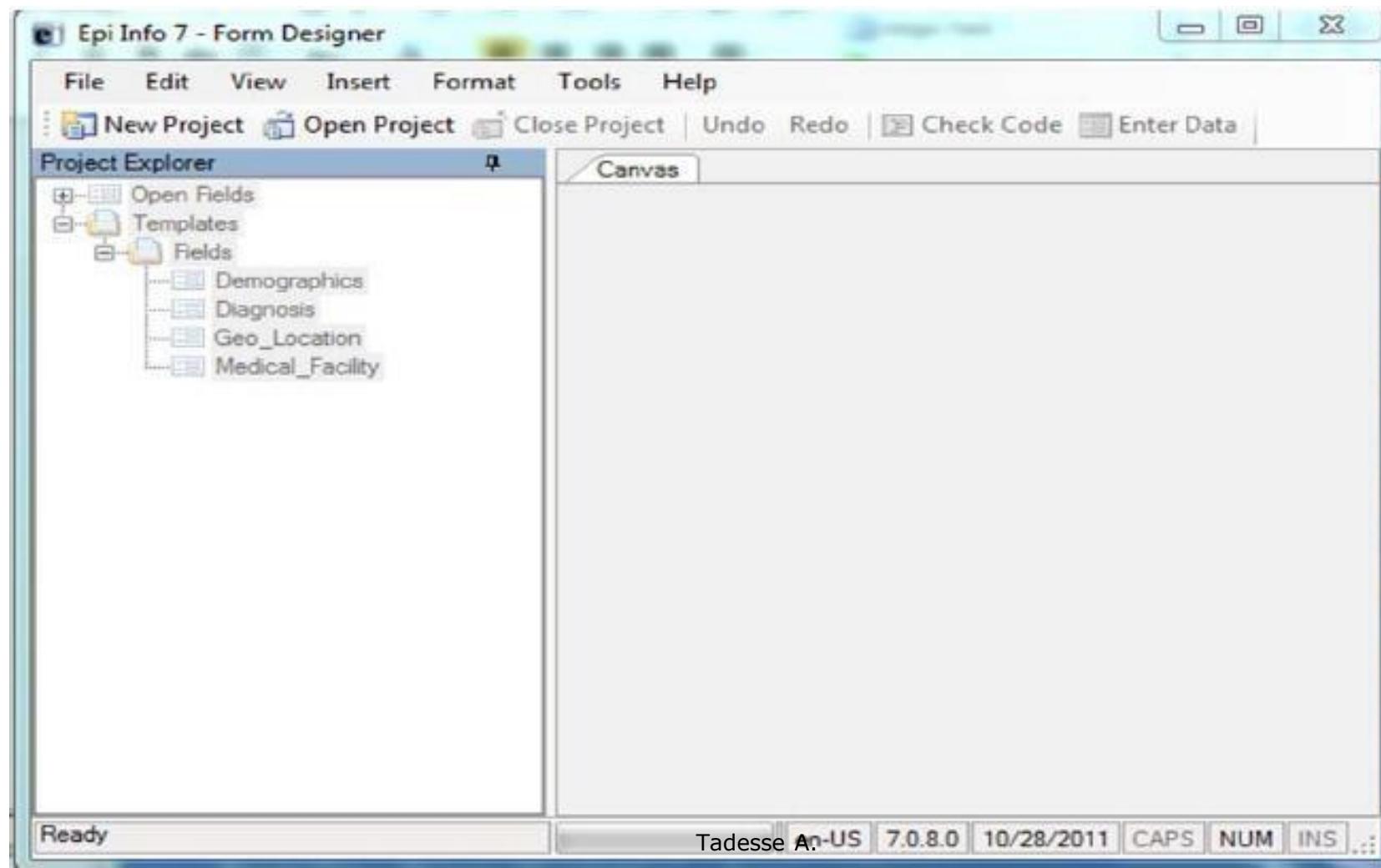
- Click «create forms» and then blank sheet will be displayed



Click create form to start

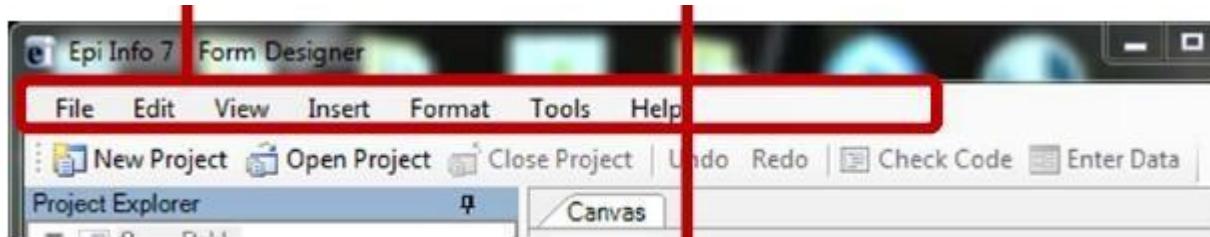
New menu will be opened  
There we can start creating forms

# Form Designer menu



# Creating forms

## Pull down menu

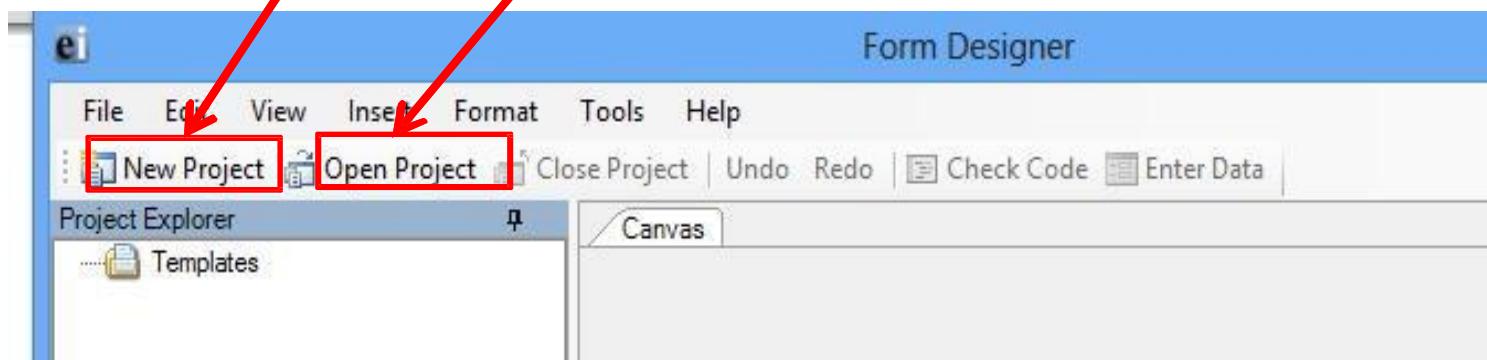


- We are able to use either the Pull down or Tool Bars menu, to create a new or open project

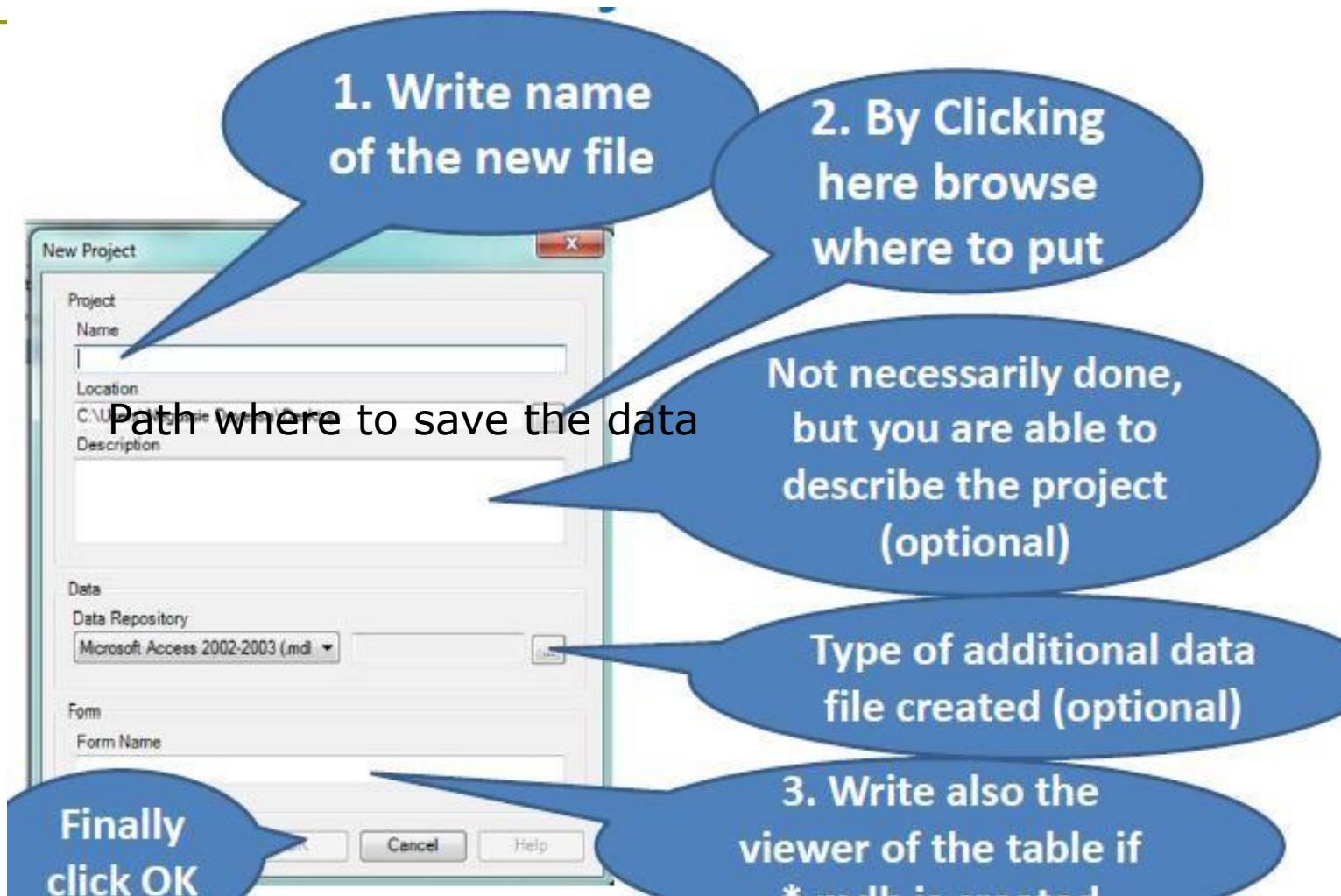
# Creating new form

---

- A new template is created by either
  1. Clicking the new project from the tool bar or
  2. Click open project to open existing form

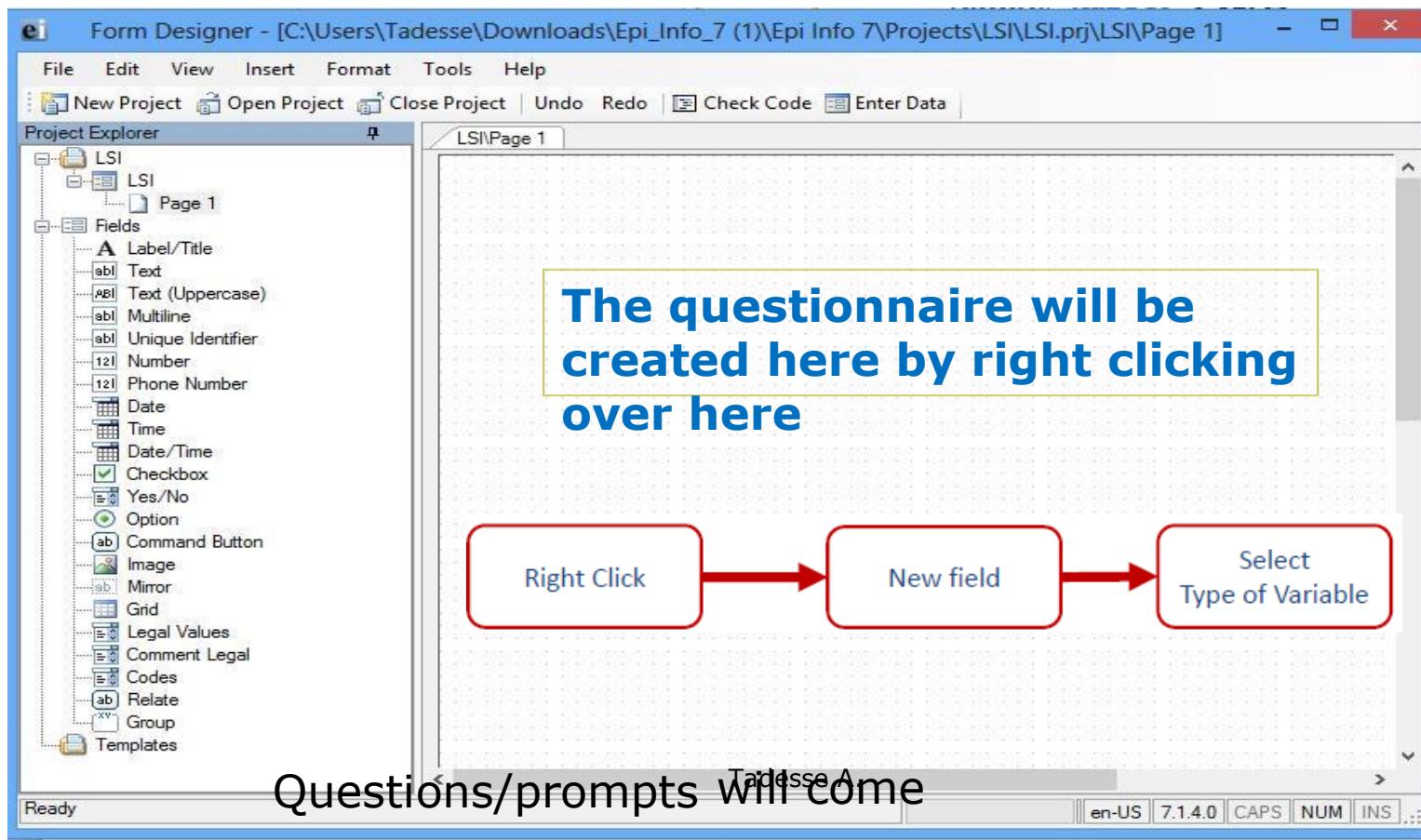


# Creating project name

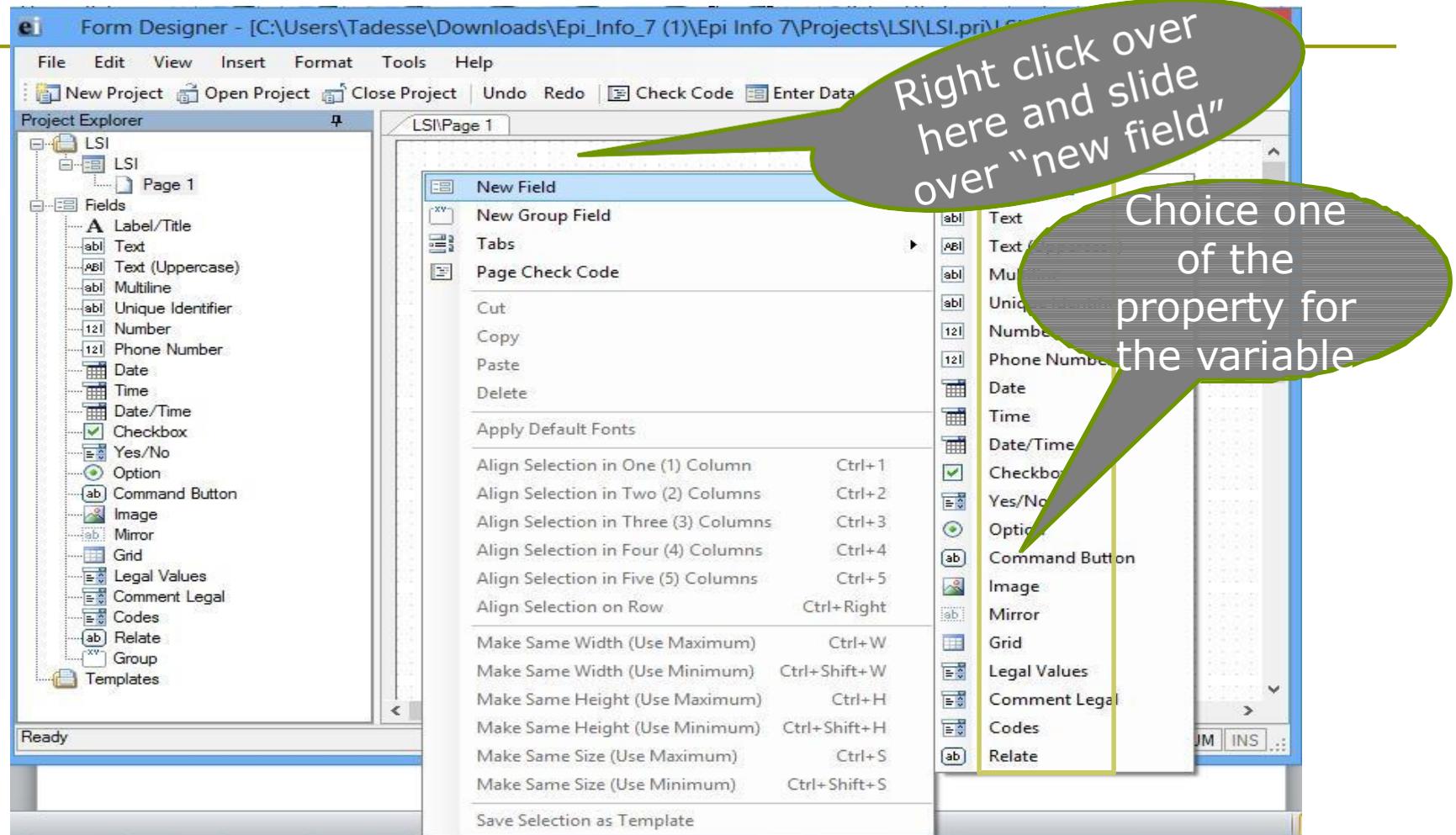


# Design menu

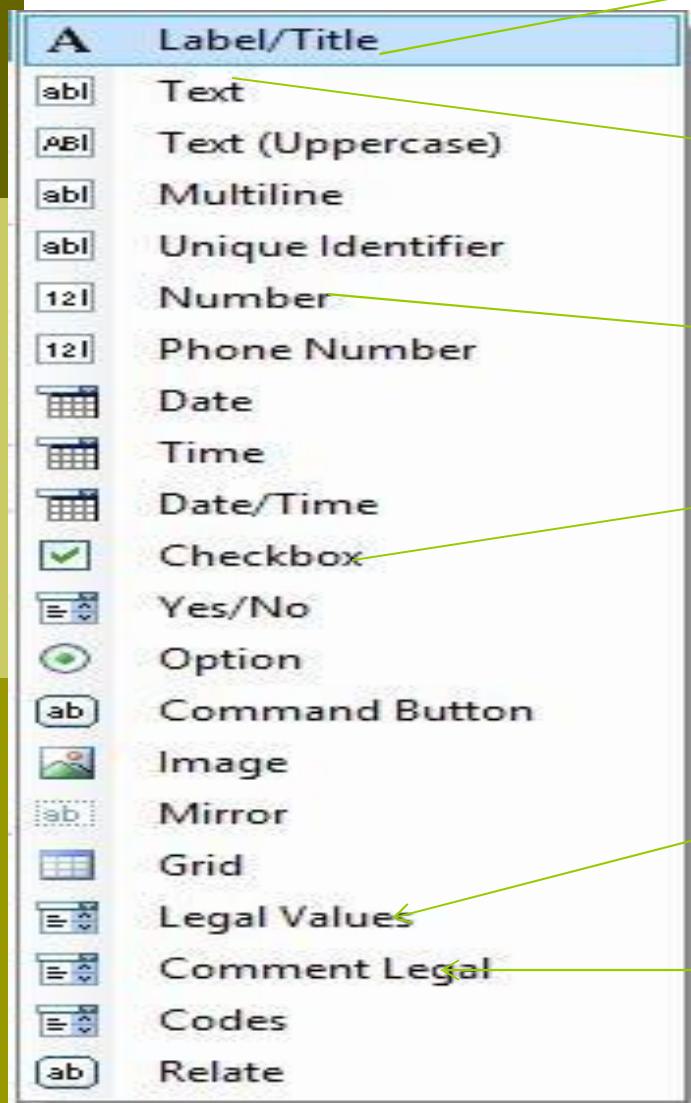
- When you create name for the project, the following box will be displayed



# Creating question/prompt



# Field types



If the variable is label or title  
(ex. For check box option,  
title the group)

**If the variable accept text**

If the variable accept numbers

If the variable allows  
multiple response

If the variable is deterministic,  
and we want the exact category to  
be displayed

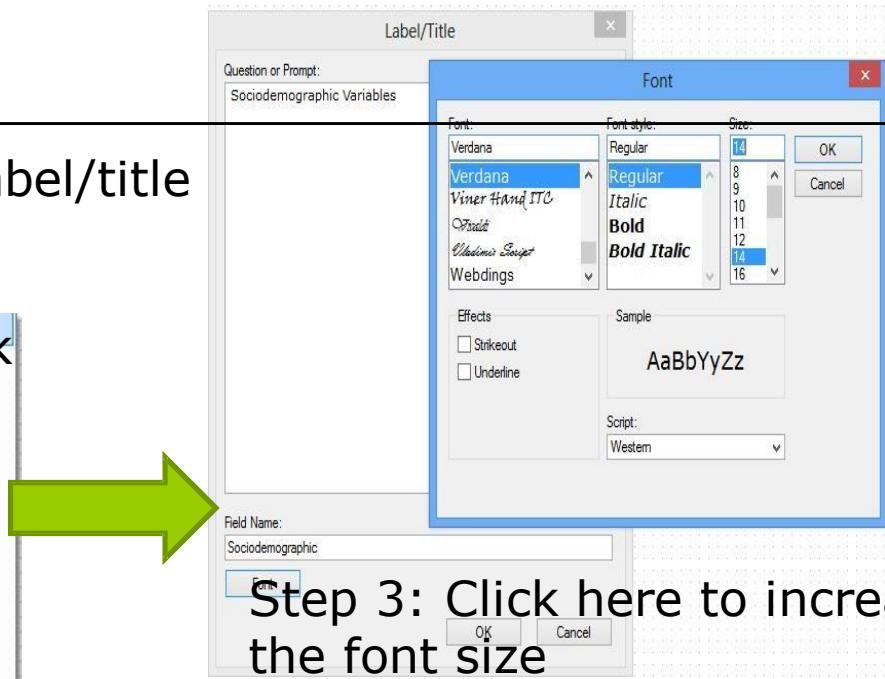
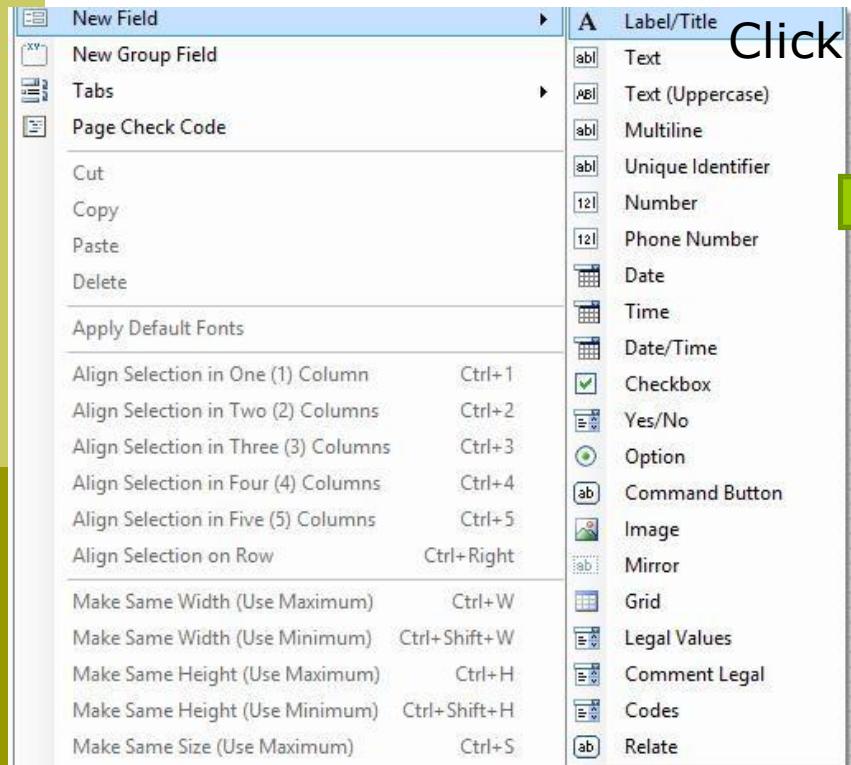
If the variable is deterministic, but  
we want numbers to refer the

Table  
Value

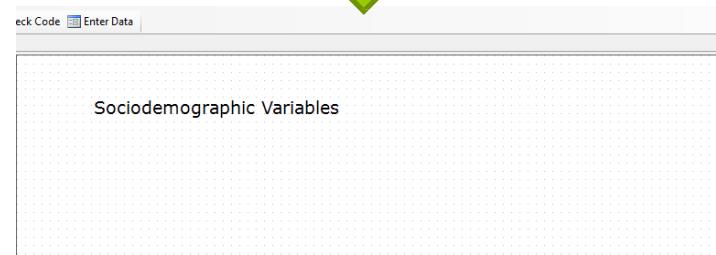
# Creating questions

Step 1: click ~~on create form~~

Step 2: under new field click label/title



Step 3: Click here to increase  
the font size



## Structure of questionnaire

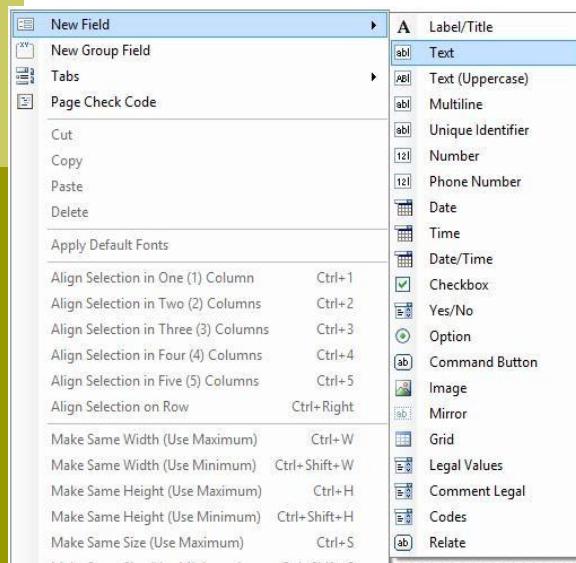
---

- Three sections:

1. Text describing field (question/ code)
2. Field name (variable coded)
3. Input definition (number/ text/ date)
4. Number of digit/ alphabet / date format

# Text describing field (question/ code)

- This is part or full description of the question
- It may be simply written the whole or part of the question
- It can also be pasted from the questionnaire
- It is a description of the coded field name



Click here to  
get variable  
name

Tadesse A.

# Text variables

- Information of text and/or numbers
  - Holding information (e.g. names, addresses, others (specify))
- Text (UPPER CASE)
  - Lower case variable automatically converted into upper case text (ex: Egypt converted into EGYPT)
- Text
  - Can hold lower case (small alphabets)
  - No mathematical operations
  - Length (How many characters)



# Field name (variable/ code name)

- ❑ It is the coded name of a variable
- ❑ Questionnaire number + few alphabets
- ❑ First word only
- ❑ No more than 10 characters
- ❑ Begin with a letter (donít begin by number)
- ❑ No spaces or punctuation marks

The variable we defined is Facility name

Tadesse A.

A screenshot of a data entry application window. At the top, there are two buttons: 'Check Code' and 'Enter Data'. Below the buttons, the text 'Sociodemographic Variables' is displayed. In the center, there is a text input field with the placeholder text 'Name of the facility'. A green arrow points from the text 'The variable we defined is Facility name' in the previous slide to this input field.

# Unique identification

Sociodemographic Variables

Name of the facility

Unique Identifier

Question or Prompt:  
Identification number

Field Name:  
ID

Attributes

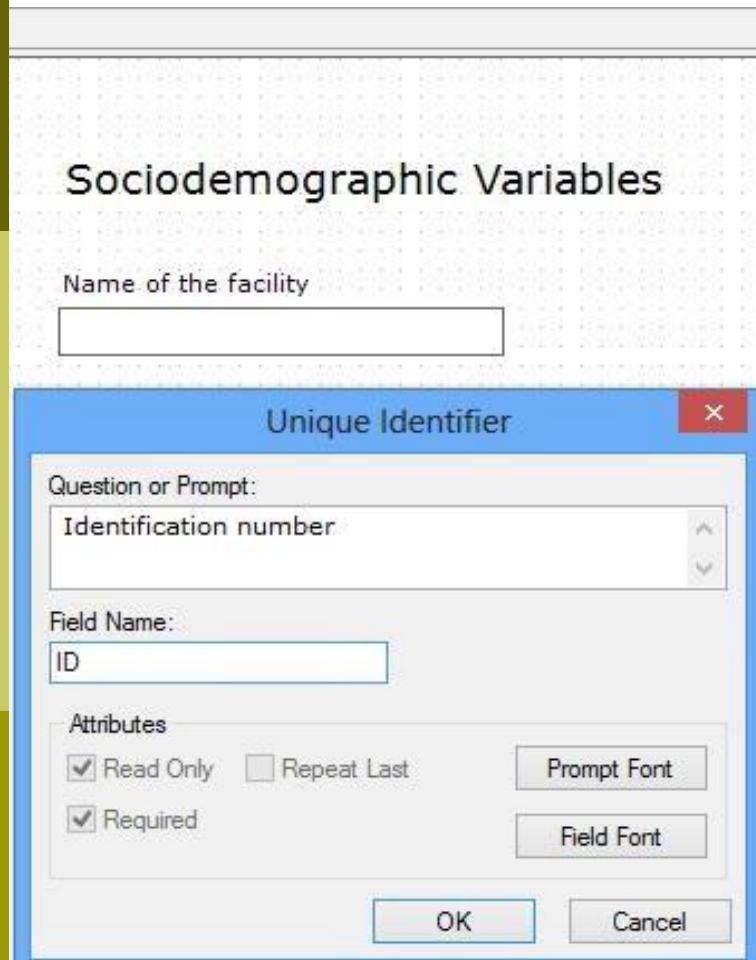
Read Only    Repeat Last

Required

Prompt Font

Field Font

OK Cancel



Check Code Enter Data

Sociodemographic Variables

Name of the facility

Identification number

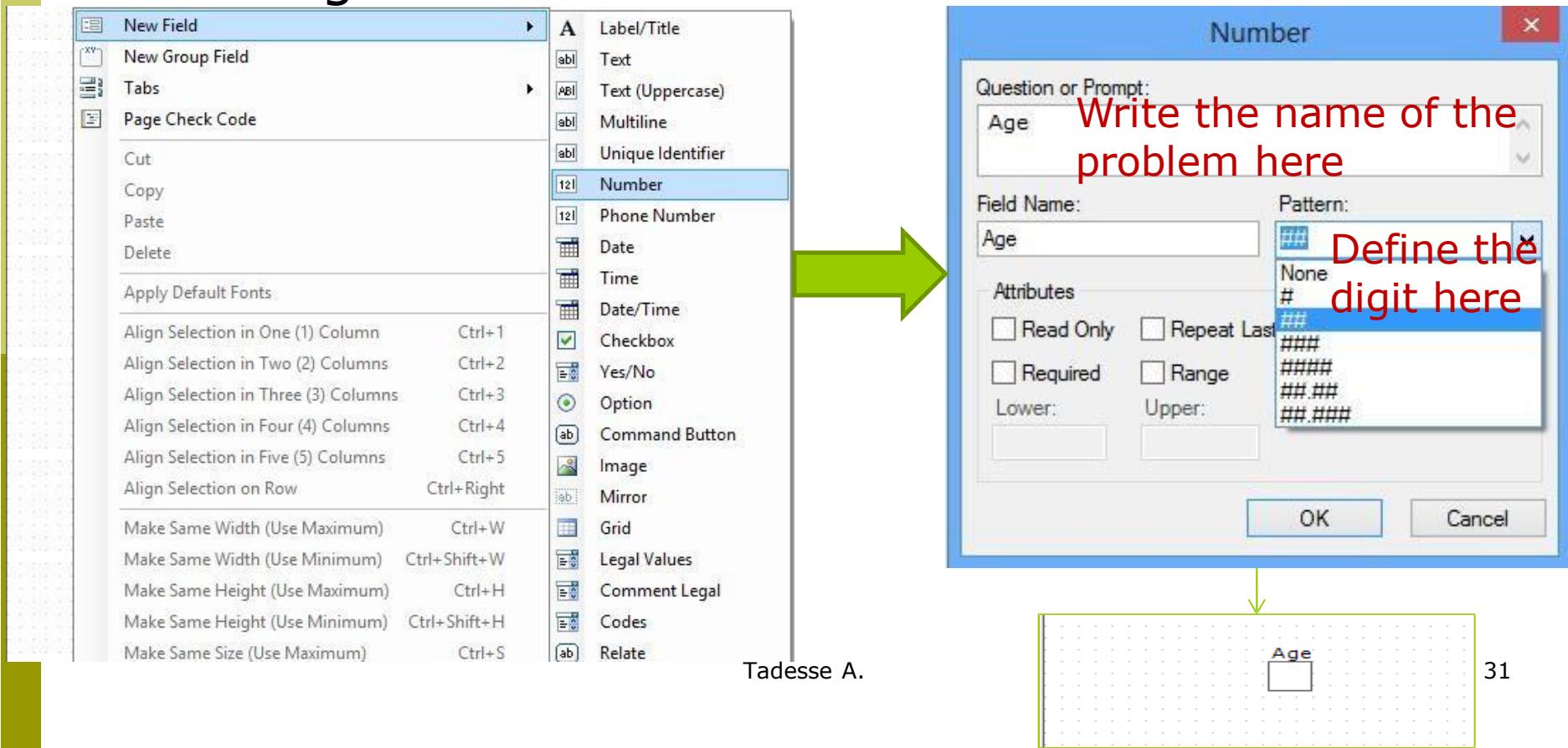
## Numeric variables (1)

---

- Continuous / discrete variables are included here
- Usually, categorical variables are coded as numerical
- Numerical information
- Holds integers (whole numbers) or numbers with a decimal point
- Length (digits, decimals after the comma)
- They are displayed with <#>, <##.#>
- You are able to type the numeric sign # manually

# Numeric variables (1)

- When the variable to be defined is a numeric variable, choose the "Number" after right clicking on the create form window



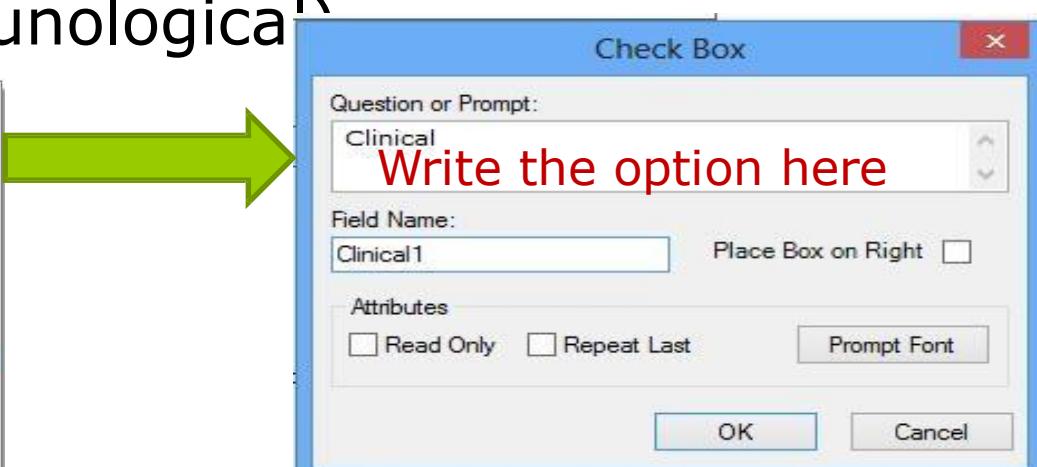
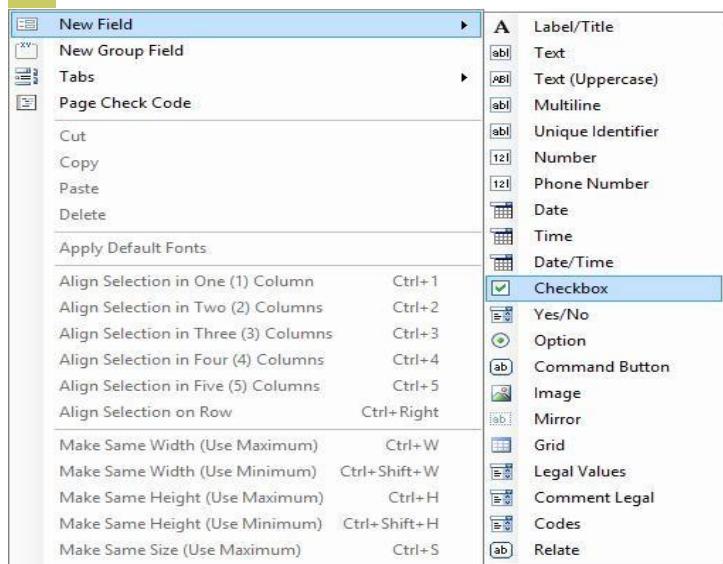
# Date variables

- Date variables:
  - Holds information on date:
    - ñ Data in american <MM/DD/YYYY>
    - ñ European <DD/MM/YYYY>



# Checkbox

- ❑ Checkbox is required when the question allows multiple response
- ❑ Example: Type of treatment failure (clinical, virological, immunological)



A screenshot of a user interface showing a question 'Type of treatment failure' enclosed in a dotted box. Below the question are three checkboxes labeled 'Clinical', 'Virological', and 'Immunological'.

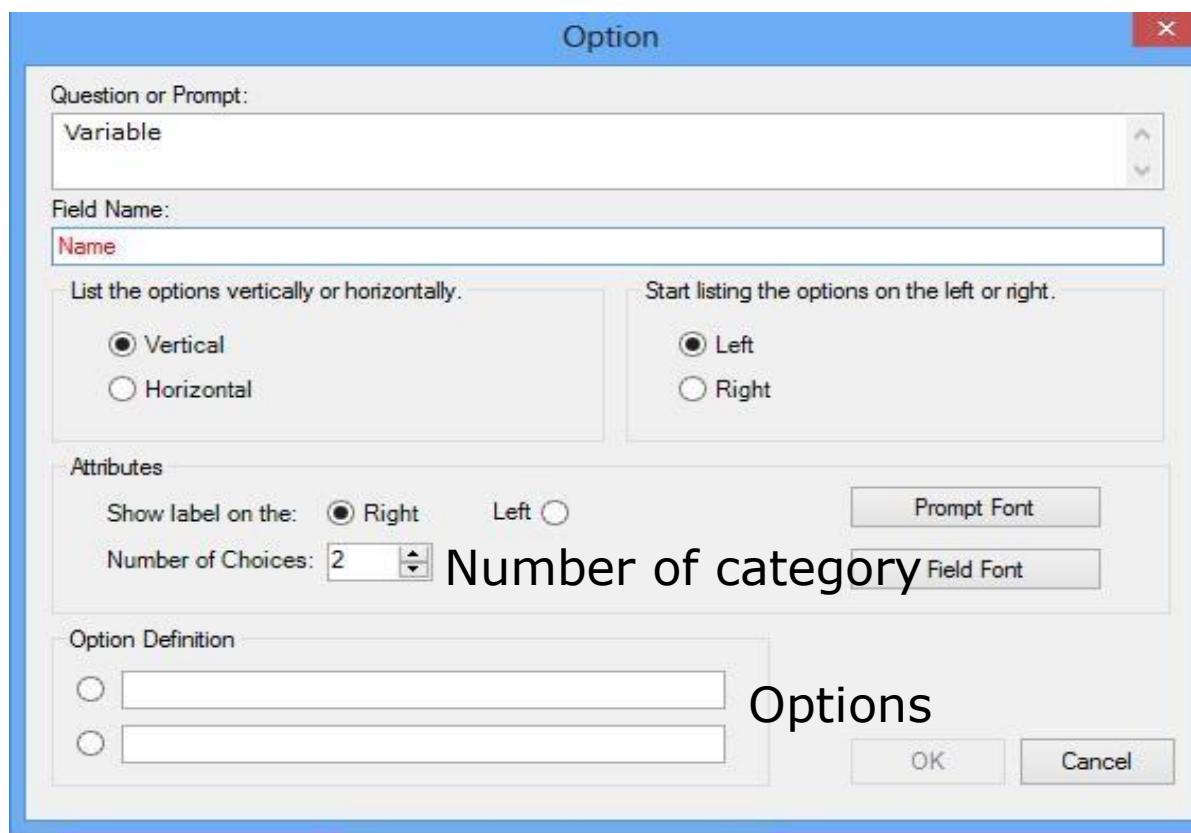
Tadesse A.

33

# Option

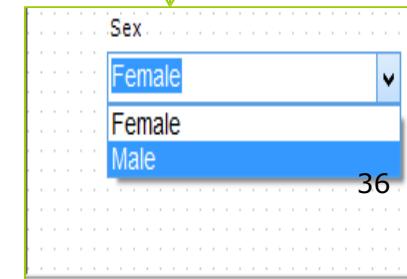
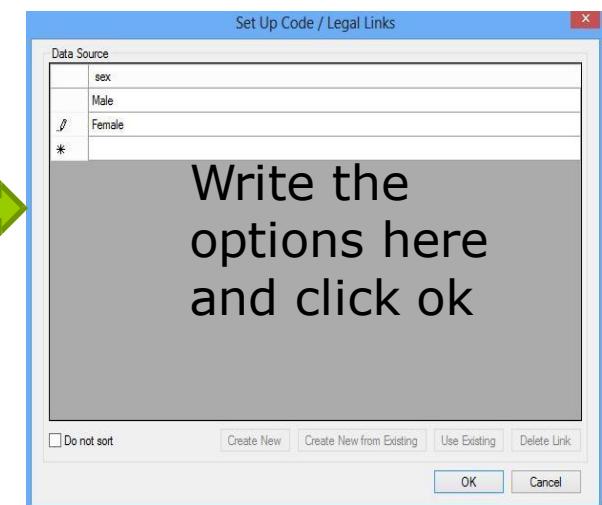
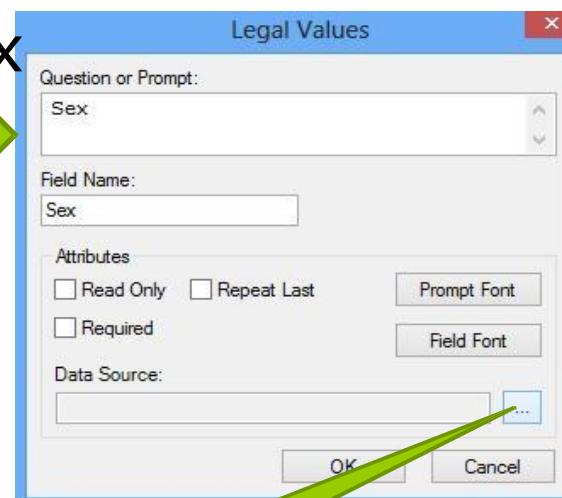
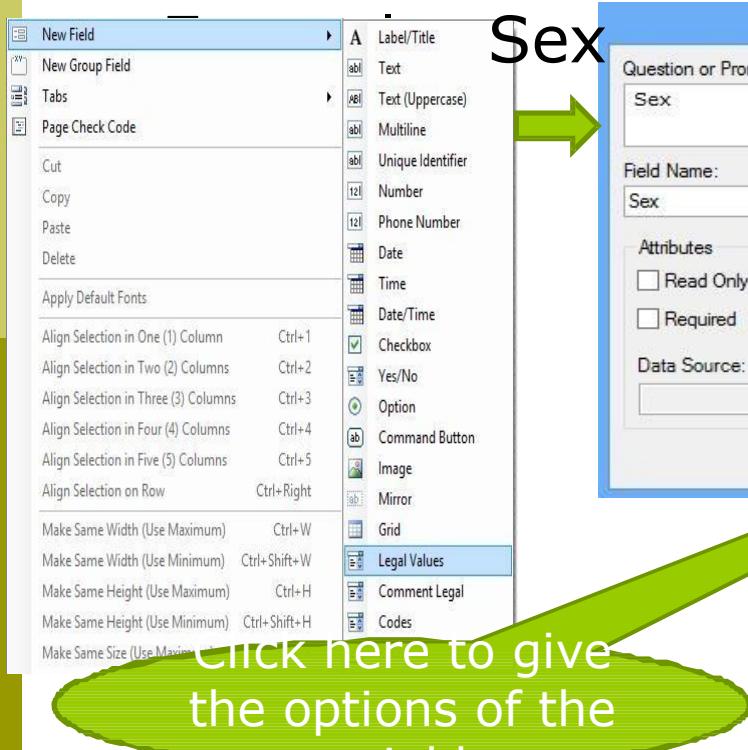
---

- ❑ This procedure can be used to define the field if the variable takes only one option



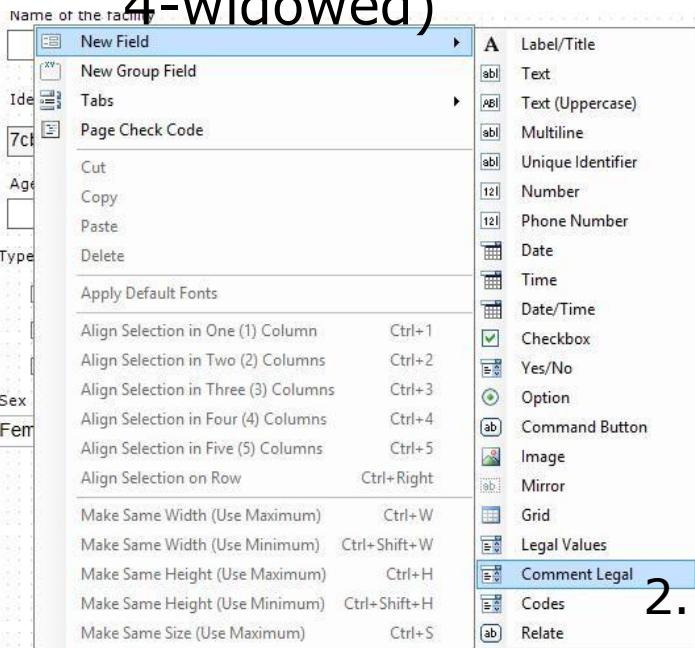
# Legal value

- When the variable is multiple choice (deterministic), exclusive and only one choice is possible, then we can define the variable using legal value

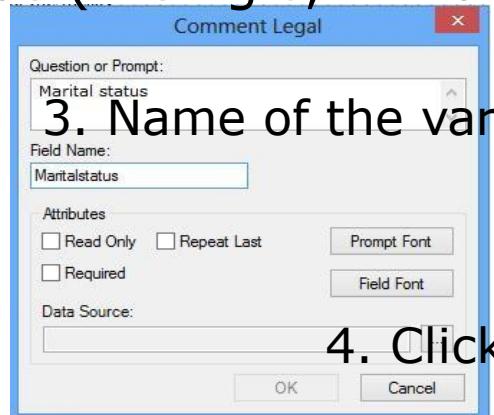


# Comment legal value

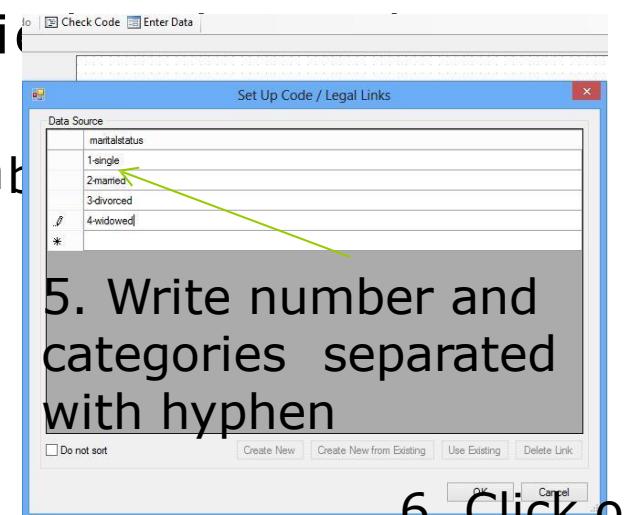
- We use comment legal value, if we need the computer to read the numerical codes instead of the real value of the variable
- Example: Marital status (1-single, 2-married, 3-divorced, 4-widowed)



2. Click

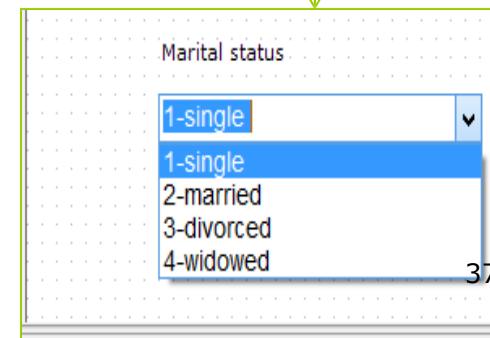


4. Click



5. Write number and categories separated with hyphen

6. Click o

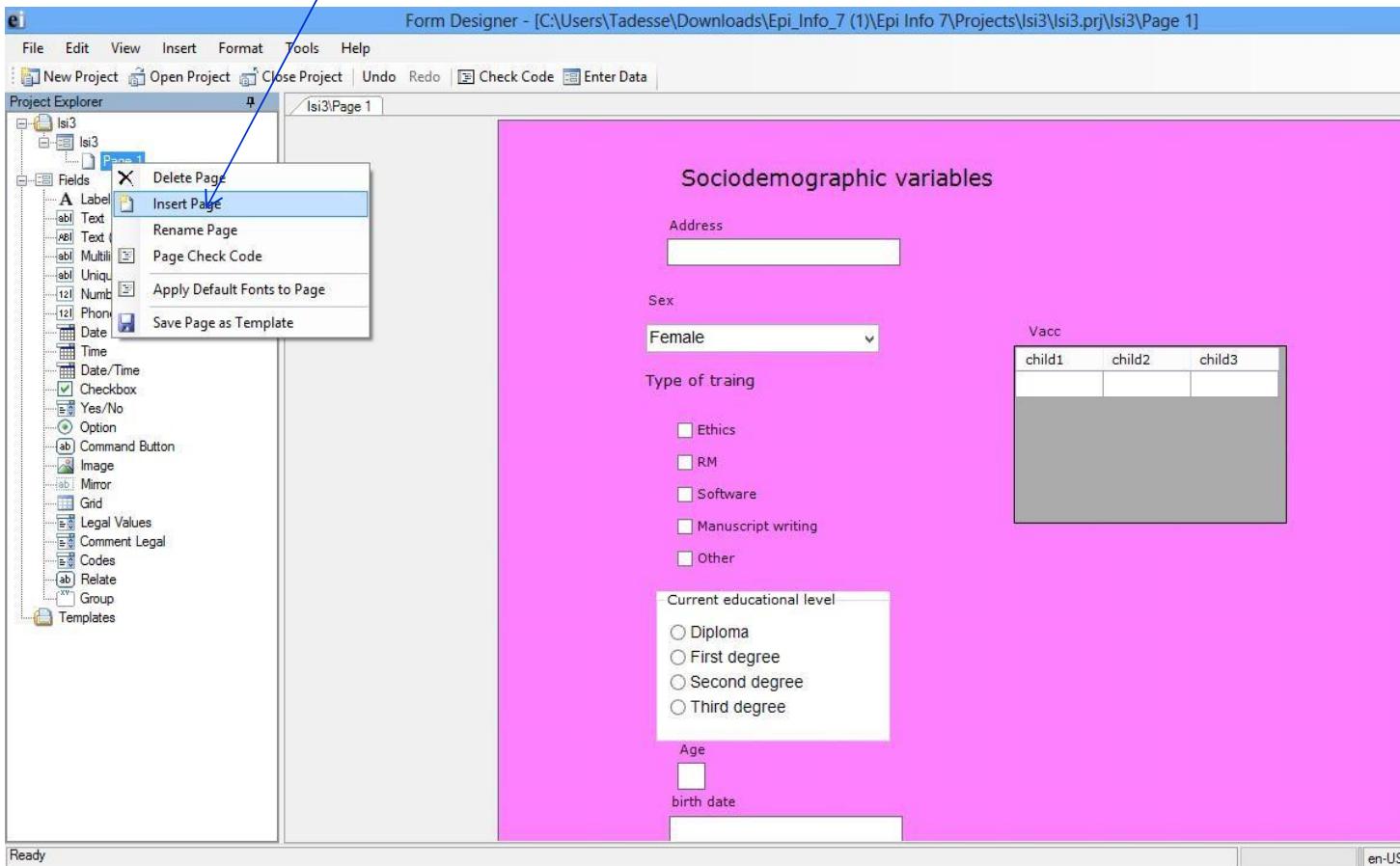


37

Tadesse A.

# Page Number

- To add next page, click on the first page and then click ï insert pageï



# Final form for questions

Form Designer - [C:\Users\Tadesse\Downloads\Epi\_Info\_7 (1)\Epi Info 7]

File Edit View Insert Format Tools Help

New Project Open Project Close Project Undo Redo Check Code Enter Data

Project Explorer LSI LSI Page 1

Sociodemographic Variables

Name of the facility

Identification number

7cb780b9-20dc

Age

Type of treatment failure

Clinical

Virological

Immunological

Sex

Female

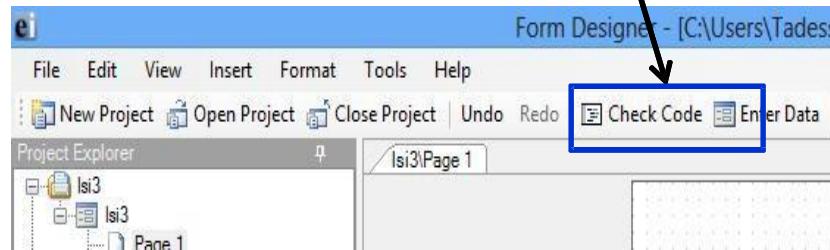
Marital status

1-single

The screenshot shows the Epi Info 7 Form Designer interface. The title bar reads "Form Designer - [C:\Users\Tadesse\Downloads\Epi\_Info\_7 (1)\Epi Info 7]". The menu bar includes File, Edit, View, Insert, Format, Tools, and Help. Below the menu is a toolbar with New Project, Open Project, Close Project, Undo, Redo, Check Code, and Enter Data buttons. The Project Explorer on the left shows a project named "LSI" containing a "LSI" folder and a "Page 1" file. The "Fields" section lists various data types: Label/Title, Text, Uppercase, Multiline, Unique Identifier, Number, Phone Number, Date, Time, Date/Time, Checkbox (selected), Yes/No, Option, Command Button, Image, Mirror, Grid, Legal Values, Comment Legal, Codes, Relate, Group, and Templates. The main workspace on the right displays a form titled "Sociodemographic Variables" with fields for Name of the facility (empty), Identification number (7cb780b9-20dc), Age (empty), Type of treatment failure (with Clinical, Virological, and Immunological options), Sex (Female), and Marital status (1-single). The bottom of the form has a footer with the text "Tadesse A.".

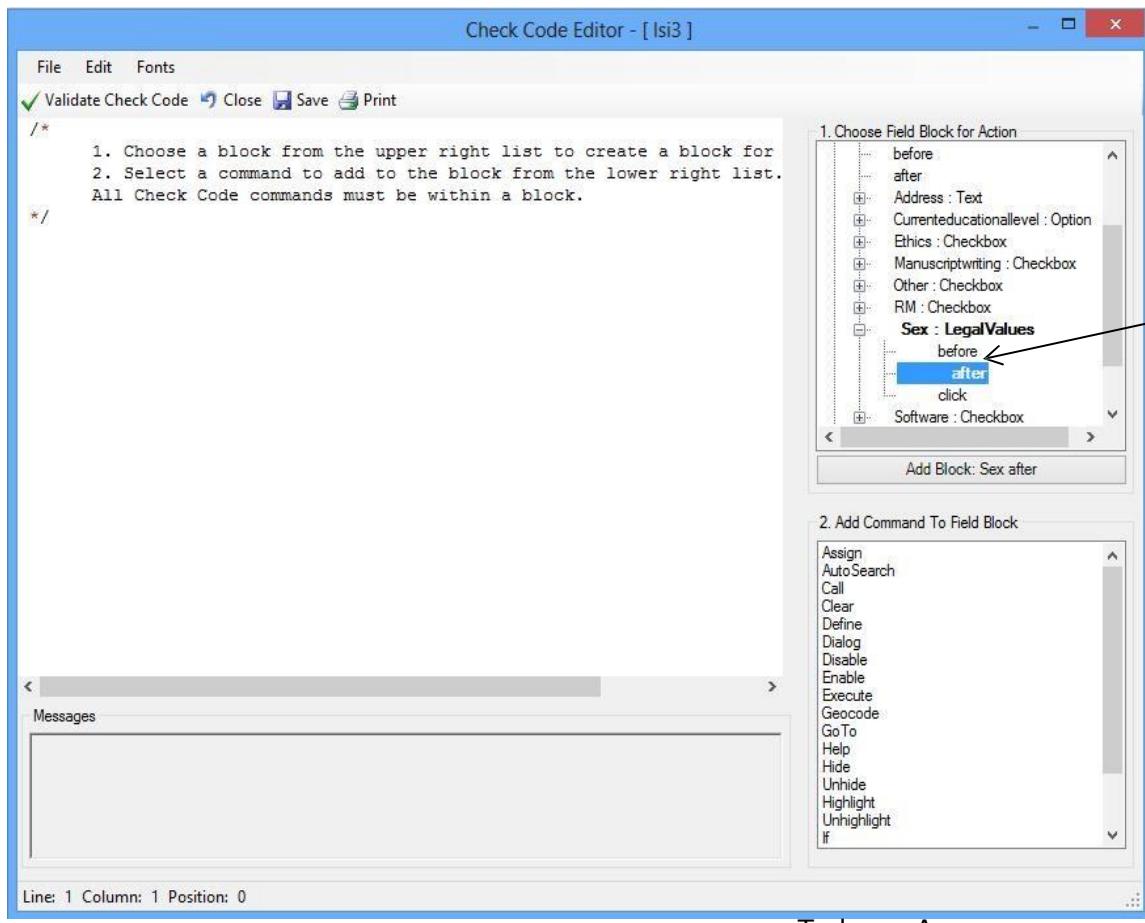
# Skip pattern

- When we administered the questionnaire, it is nonsense to ask the respondent about the frequency of smoking if he/she replayed No for the question ï Have you ever smoke?î
- This procedure helps us to jump to the question which the respondents are not legible
- To apply this application, click ï **check code** î in the main menu bar



# Skip pattern

- When you click **check code** you will get the following;



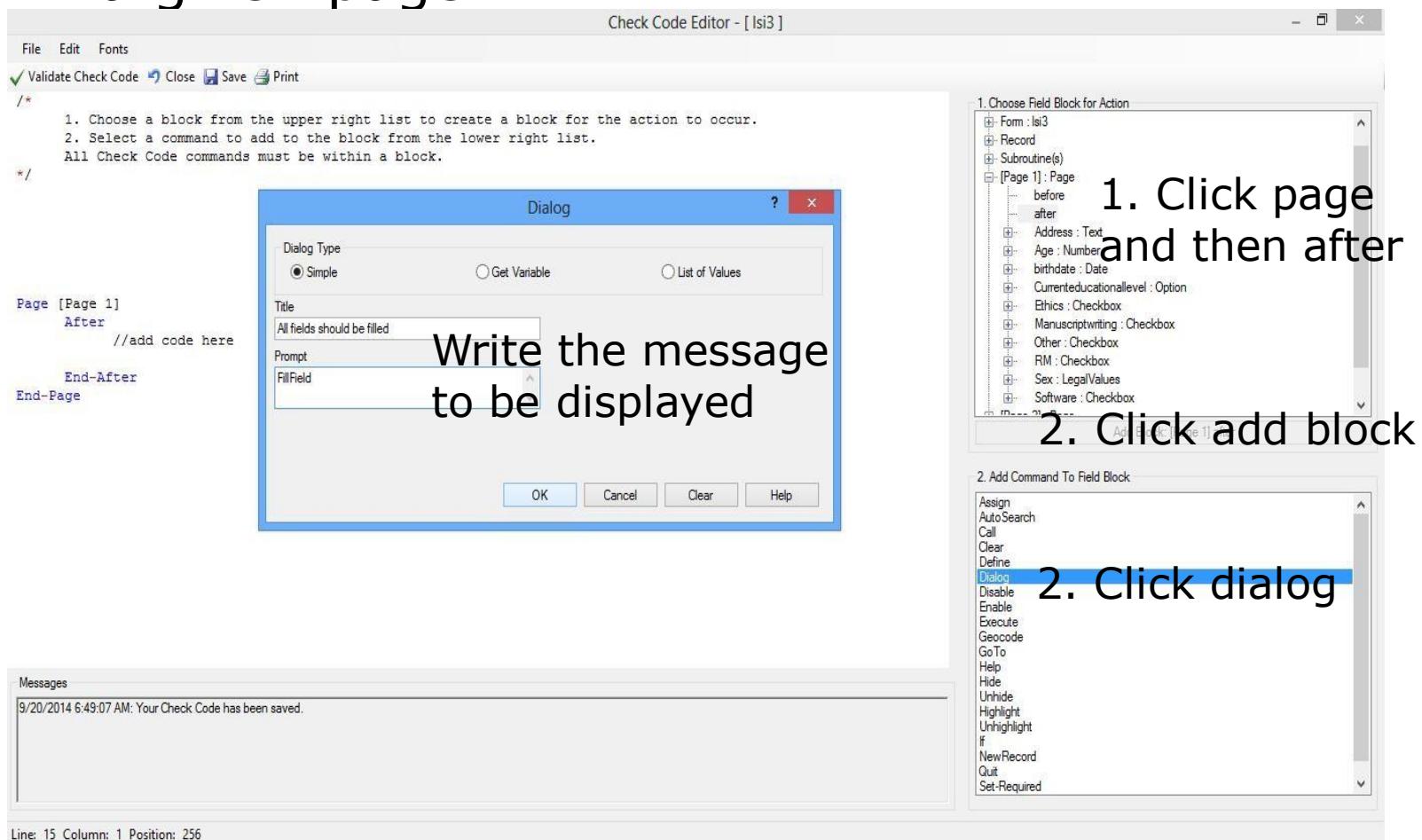
Variable of which the condition will be set

Click **add block**

Condition to be defined

# Skip pattern

- Make restriction for not skipping a single question in a given page



# Skip pattern

---

- We can also generate new variables from existing variables
- Example: Generate age from date of birth
  - Step 1. click ì check codeî
  - Step 2. click page where the variable is there
  - Step 3: click ì date of birthî
  - Step 4: click ì add block..î
  - Step 5: click ì assignî
  - Step 6: Choose the variable age
  - Step 7: under the ì expressionî , write years and then open bracket
  - Step 8: click the variable ì date of birthî under **available variable**

# Skip pattern

---

- Step 9: Click the lift side of expression and choose system function and then ìSYSTEMDATEî in order to use current date.
- Otherwise, define the current date
- Finally close the bracket and click ok

# Creating age from date of birth(1)

Check Code Editor - [ lsj3 ]

File Edit Fonts  
✓ Validate Check Code Close Save Print

```
/*  
1. Choose a block from the upper right list to create a block for the action to occur.  
2. Select a command to add to the block from the lower right list.  
All Check Code commands must be within a block.  
*/  
  
Page [Page 1]  
    After  
        //add code here  
        DIALOG "FillField" TITLETEXT="All fields should be filled"  
  
    End-After  
End-Page  
  
Field birthdate  
    After  
        //add code here  
  
    End-After  
End-Field
```

Assign

Choose age here

Assign Variable: Age

= Expression: years (birthdate,

Available Variables: birthdate

Write 1 years!

OK Cancel Functions Clear Help

Date of birth and then after

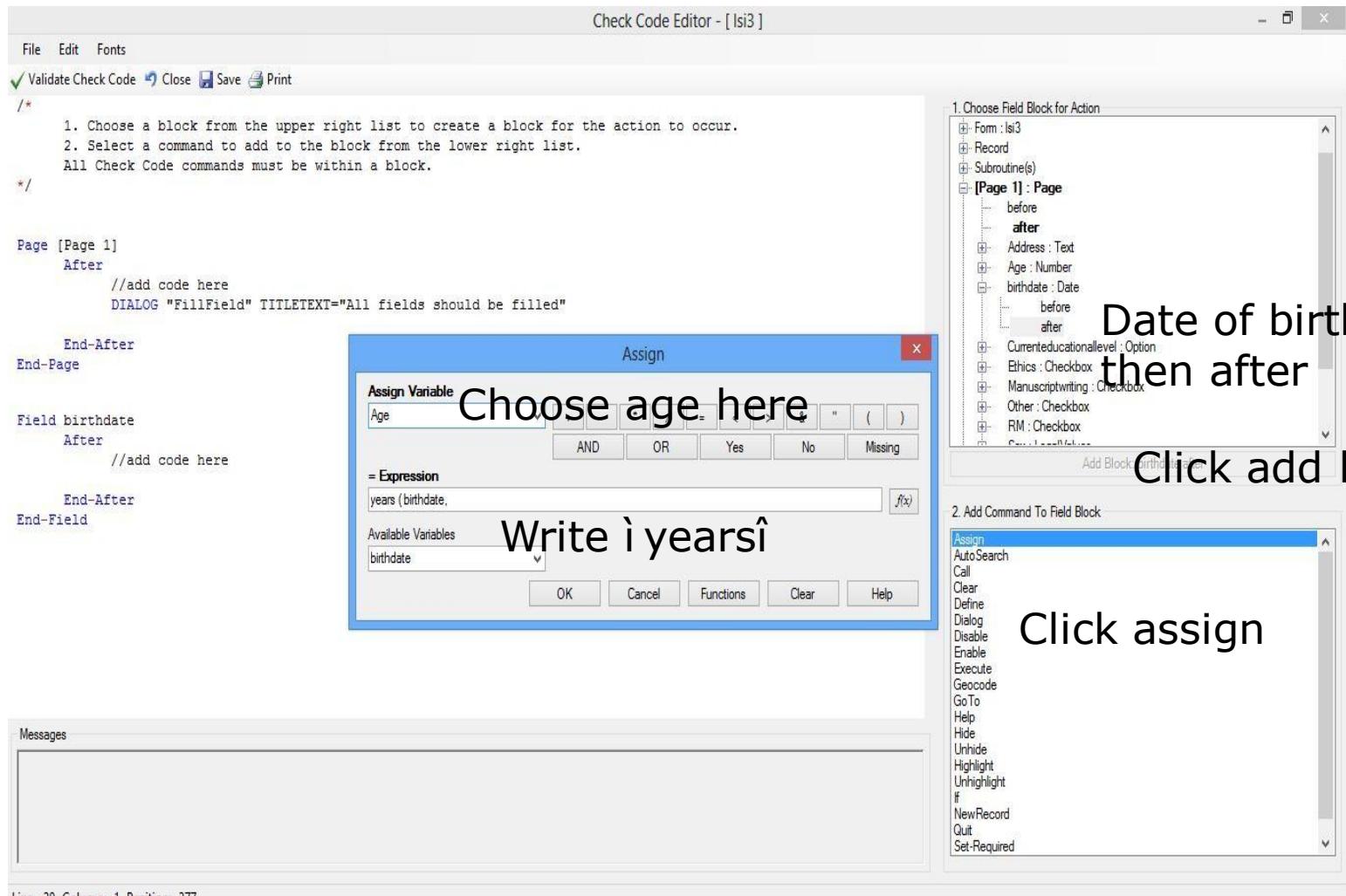
Click add block

Click assign

Messages

Line: 20 Column: 1 Position: 377

adresse A.



The screenshot shows the 'Check Code Editor' window with the title 'Check Code Editor - [ lsj3 ]'. The menu bar includes 'File', 'Edit', 'Fonts', 'Validate Check Code', 'Close', 'Save', and 'Print'. A status bar at the bottom shows 'Line: 20 Column: 1 Position: 377' and 'adresse A.'.

The main code area contains:

```
/*  
1. Choose a block from the upper right list to create a block for the action to occur.  
2. Select a command to add to the block from the lower right list.  
All Check Code commands must be within a block.  
*/  
  
Page [Page 1]  
    After  
        //add code here  
        DIALOG "FillField" TITLETEXT="All fields should be filled"  
  
    End-After  
End-Page  
  
Field birthdate  
    After  
        //add code here  
  
    End-After  
End-Field
```

A modal dialog box titled 'Assign' is open in the center. It has a title bar 'Assign', a message 'Choose age here', and a text input field containing 'Age'. Below it is another text input field labeled '= Expression' with the value 'years (birthdate,'. A dropdown menu 'Available Variables' shows 'birthdate'. At the bottom of the dialog are buttons for 'OK', 'Cancel', 'Functions', 'Clear', and 'Help'.

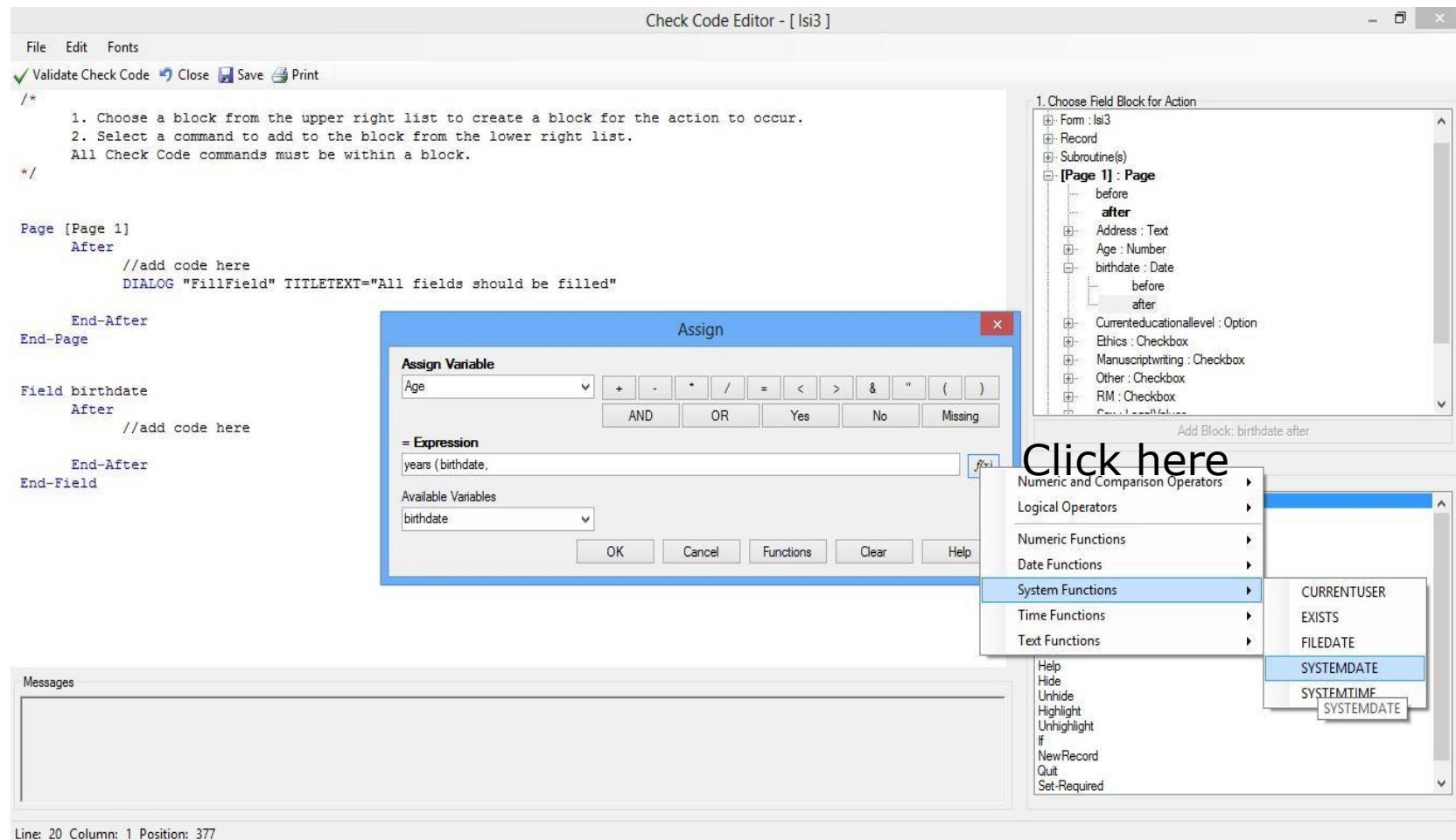
To the right of the editor, there are two vertical lists:

- The top list, titled '1. Choose Field Block for Action', shows a tree structure:
  - Form : lsj3
  - Record
  - Subroutine(s)
  - [Page 1] : Page
    - before
    - after
    - Address : Text
    - Age : Number
    - birthdate : Date
      - before
      - after
    - Currenteducationlevel : Option
    - Ethics : Checkbox
    - Manuscriptwriting : Checkbox
    - Other : Checkbox
    - RM : Checkbox
- The bottom list, titled '2. Add Command To Field Block', shows a list of commands:
  - Assign
  - Auto Search
  - Call
  - Clear
  - Define
  - Dialog
  - Disable
  - Enable
  - Execute
  - Geocode
  - Go To
  - Help
  - Hide
  - Unhide
  - Highlight
  - Unhighlight
  - If
  - New Record
  - Quit
  - Set Required

Annotations on the right side of the editor area include:

- 'Date of birth and then after' pointing to the 'birthdate' variable in the code.
- 'Click add block' pointing to the 'Add Block' button in the '1. Choose Field Block for Action' list.
- 'Click assign' pointing to the 'Assign' command in the '2. Add Command To Field Block' list.

# Creating age from date of birth (2)



# Final codes

Check Code Editor - [ lsi3 ]

File Edit Fonts  
✓ Validate Check Code Close Save Print

```
/*
 1. Choose a block from the upper right list to create a block for the action to occur.
 2. Select a command to add to the block from the lower right list.
 All Check Code commands must be within a block.
 */

Page [Page 1]
  After
    //add code here
    DIALOG "FillField" TITLETEXT="All fields should be filled"

  End-After
End-Page

Field birthdate
  After
    //add code here
    ASSIGN Age = years ( birthdate, SYSDATE )

  End-After
End-Field
```

Page restriction

Age generation

Finally save the changes you made

Messages

1. Choose Field Block for Action

- + Form : lsi3
- + Record
- + Subroutine(s)
- [Page 1] : Page
  - before
  - after
    - + Address : Text
    - + Age : Number
    - + birthdate : Date
      - before
      - after
    - + Currenteducationlevel : Option
    - + Ethics : Checkbox
    - + Manuscriptwriting : Checkbox
    - + Other : Checkbox
    - + RM : Checkbox
    - + ...

Add Block: birthdate after

2. Add Command To Field Block

- Assign
- AutoSearch
- Call
- Clear
- Define
- Dialog
- Disable
- Enable
- Execute
- Geocode
- GoTo
- Help
- Hide
- Unhide
- Highlight
- Unhighlight
- If
- NewRecord
- Quit
- Set-Required

Line: 20 Column: 1 Position: 377

adresse A.

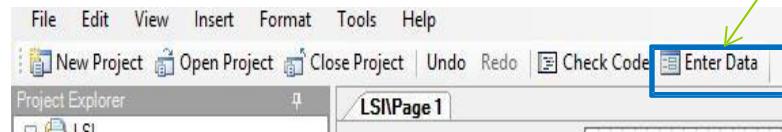
## Exercises 1: Design template for the following questions

---

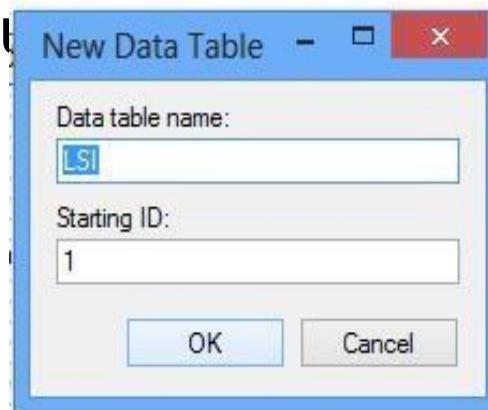
S. No	Question	Category /Answer
1	ID	
2	Field of study	1. Health                            2. Natural science 3. Social science                    4. Business and economics 5. Others
3	Sex	1. Male    2. Female
4	Age at first formal school	1. Age in years_____
5	Family residence	1. Urban                            2. rural 3. semi-urban
6	Current education level	1. First degree                    2. Secondary degree 3. Other
7	Have you ever participated in training like this as a trainee before ?	1. yes 2. No
8	Type of training	1. Ethics in health research      2. Research Methods 3. Statistical software            4. Teaching methodology 5. Manuscript writing             6. other , specify
9	How do you level the added knowledge of EPI-info session of this training to you?	1.Excellent                        2. Good 3. little                             4. Not at all

# Data entry

- When you complete the design, click **Enter data** to open data entry form



- It asks you to create data table and click ok
- Then give name for the data table. To use the same name as your form, click ok



# Data entry

- Here is the data entry form

The screenshot shows a software interface for data entry. At the top, there's a toolbar with various icons: Open Form, Save, Print, Find, New Record, etc. Below the toolbar, the title 'Page 1' is displayed above a section titled 'Sociodemographic Variables'. This section contains several input fields:

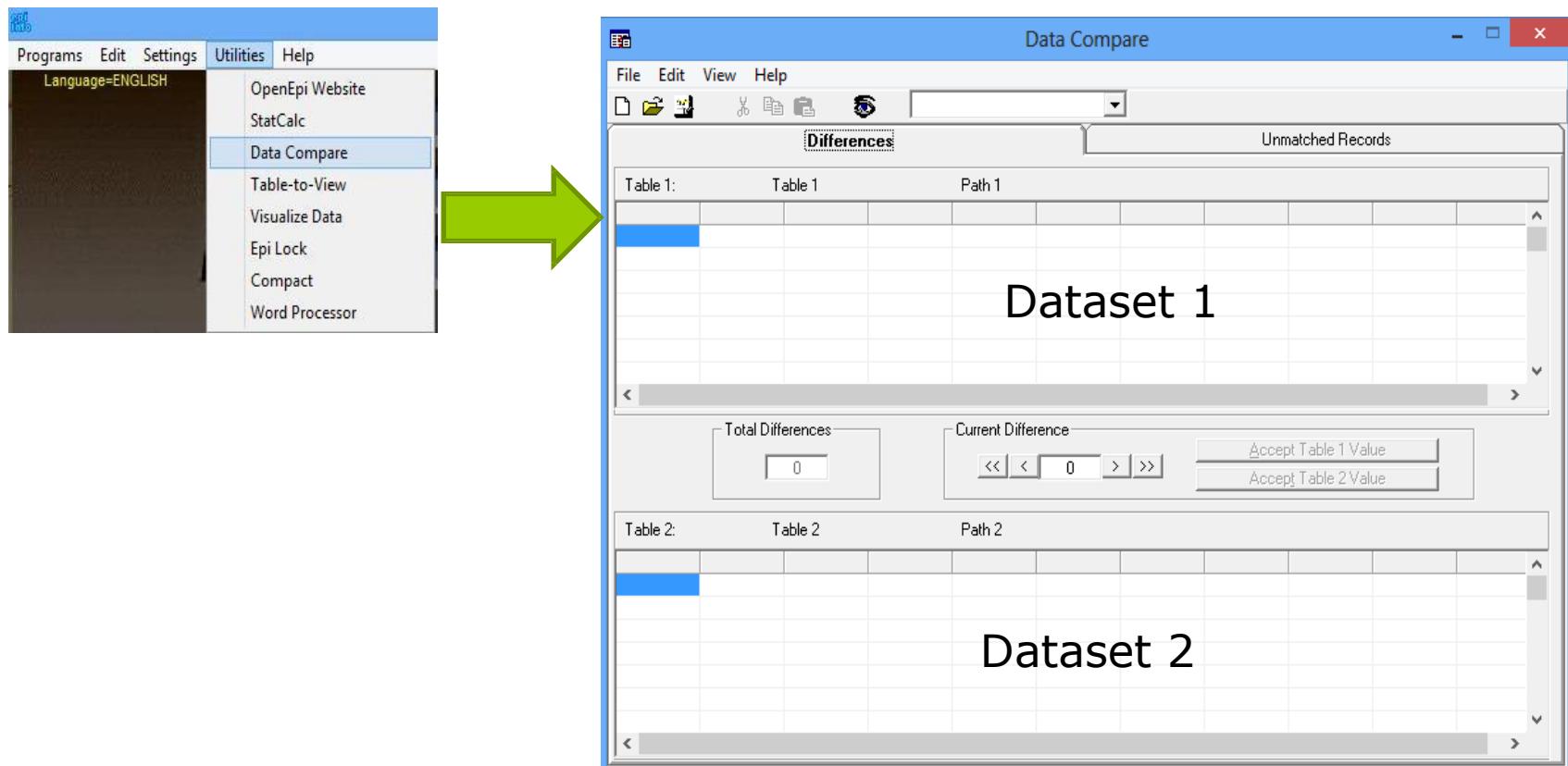
- Name of the facility: A text input field containing '1'.
- Age: A text input field containing '22'.
- Type of treatment failure: A group of three checkboxes, all of which are checked:
  - Clinical
  - Virological
  - Immunological
- Sex: A dropdown menu showing 'Male'.
- Marital status: A dropdown menu showing '2-married'.

On the left side of the main window, there's a sidebar titled 'Pages' which lists 'LSI' and 'Page 1'. Below this is a 'Linked Records' panel with tabs for 'Exposed From' and 'Exposed To', each showing a list of items. At the bottom of the sidebar, there are buttons for 'Unlink' and 'Add Exposure...', and a link to 'View SNA Graph...'. The overall interface is clean and organized, designed for efficient data collection.

- Thus, we can entry data here

# Data comparison

- To compare two datasets if they are identical, use the following procedure (EPI-Info 2002)



## Exercises 2:

---

- Communicate 5 of your colleagues
- Collect data from your colleagues
- Enter the data in to the form you created

# Other applications

---

## □ StatCalc

- Which contains:
  - ▣ Sample size and power, for
    - Population survey
    - Cohort or analytic cross-sectional design
    - Unmatched case control
  - Chi square for trend
  - Tables (2 by 2, 2 by n)
  - Poisson (rare event vs std)
  - Binomial (proportion vs std)
  - Matched pair case control study

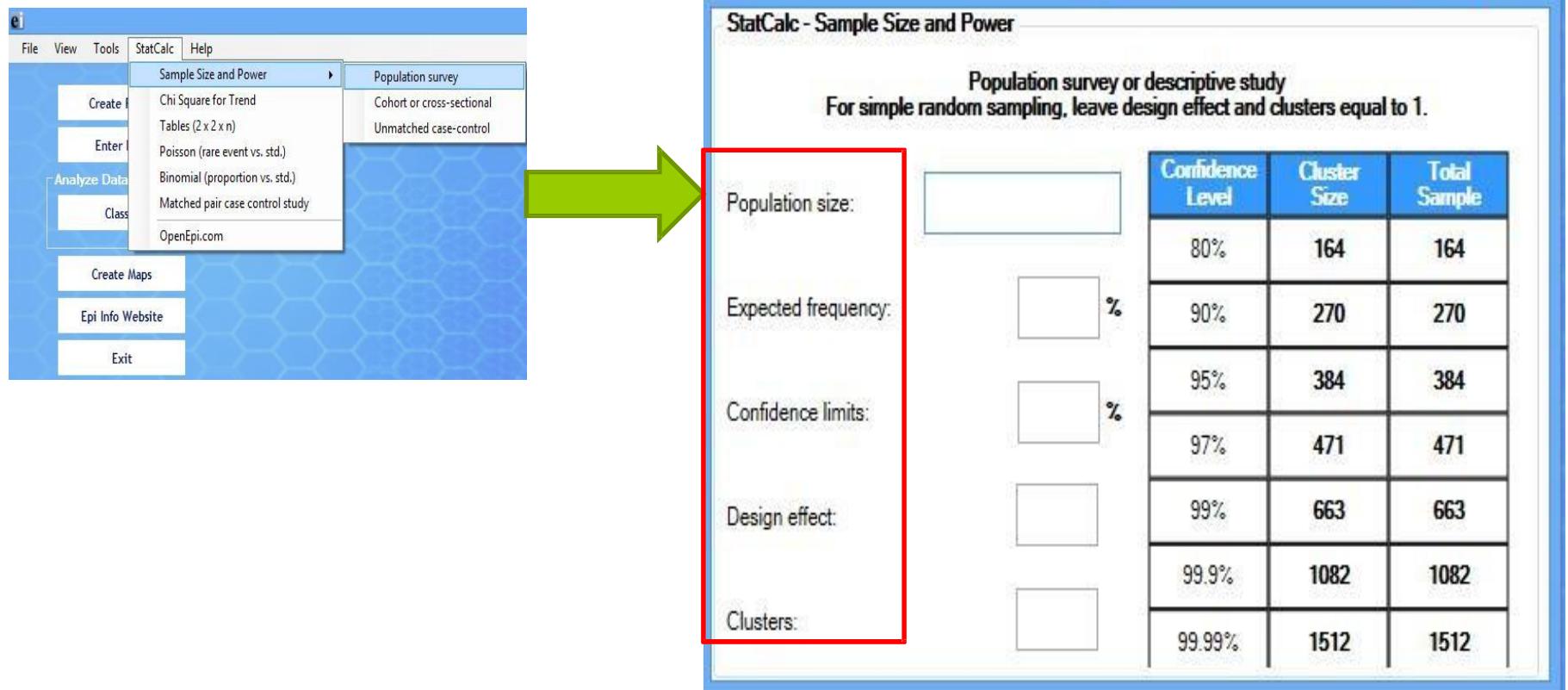
# Sample size and power

---

- Parameters needed to determine sample size for population survey are:
  - Total population (study population)
  - Expected frequency (prevalence of the variable of interest)
  - Where shall we get the value?
  - Confidence limit (worst acceptable difference, (0, 5])
  - Design effect, if needed
  - Clusters (depending on the number of clusters, you can adjust the sample size)

# Sample size and power

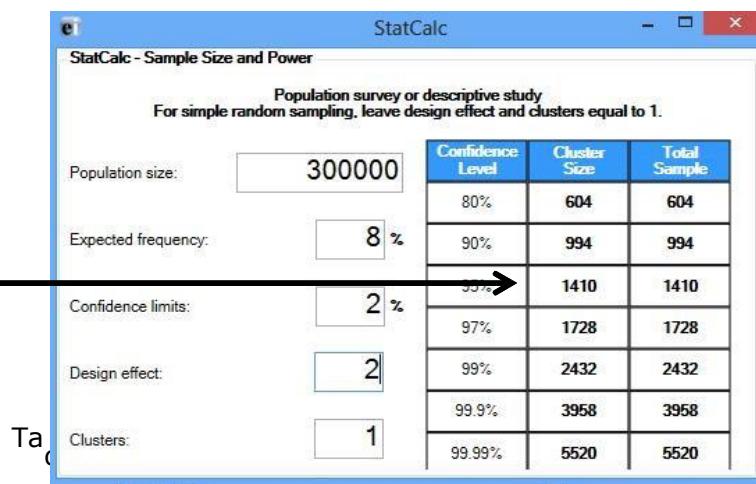
## □ Steps



# Sample size and power

- Example: Consider the following scenario
- Objective: determine the prevalence of diabetics in Gondar
  - Total population 300,000
  - Prevalence from study in Addis 8%
  - Sampling method multistage
  - Cluster 1

Sample size



Ta

# Sample size and power

## □ Cohort or analytic cross-sectional design

StatCalc - Sample Size and Power

Unmatched Cohort and Cross-Sectional Studies (Exposed and Nonexposed)

Two-sided confidence level:

Power:  %

Ratio (Unexposed : Exposed):

% outcome in unexposed group:  %

Risk ratio:

Odds ratio:

% outcome in exposed group:  %

	Kelsey	Fleiss	Fleiss w/ CC
Exposed	0	0	0
Unexposed	0	0	0
Total	0	0	0

Tadesse A.

56

# Sample size and power

## □ Unmatched case control

StatCalc - Sample Size and Power

Unmatched Case-Control Study (Comparison of ILL and NOT ILL)

Two-sided confidence level:

Power:  %

Ratio of controls to cases:

Percent of controls exposed:  %

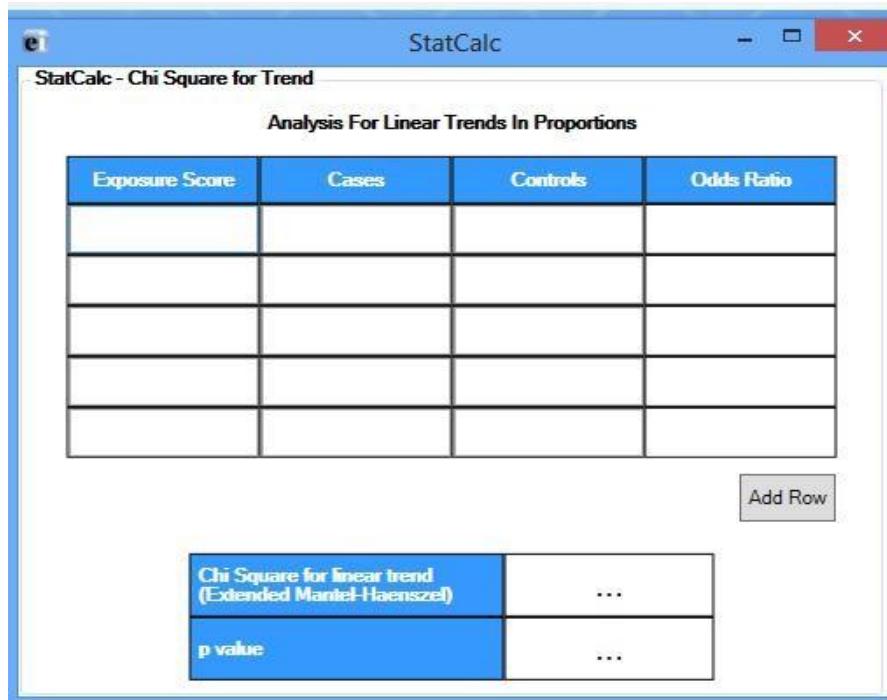
Odds ratio:

Percent of cases with exposure:  %

	Kelsey	Fleiss	Fleiss w/ CC
Cases	0	0	0
Controls	0	0	0
Total	0	0	0

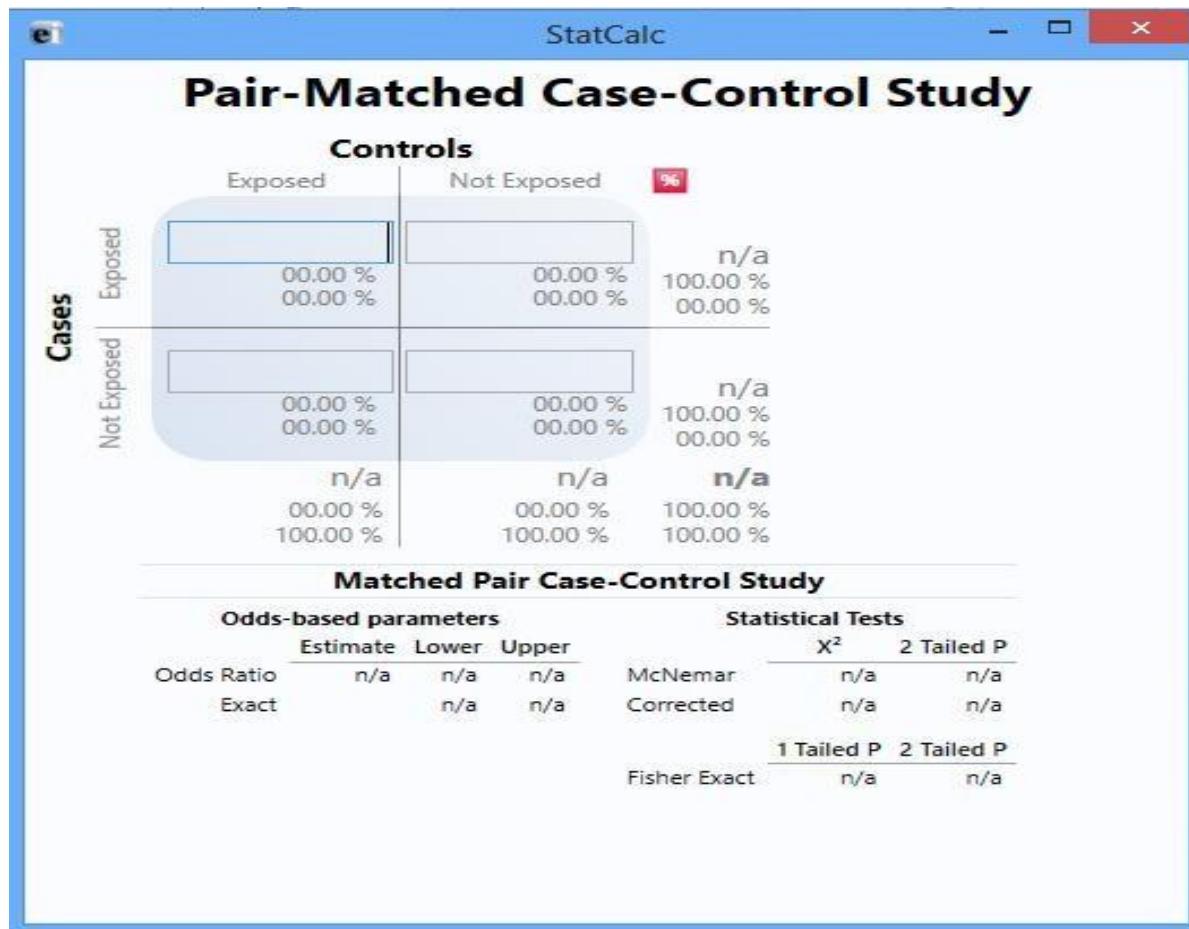
# Chi-square for trend

- If the exposure variable is ordinal, we used chi-square for trend to determine the presence of association



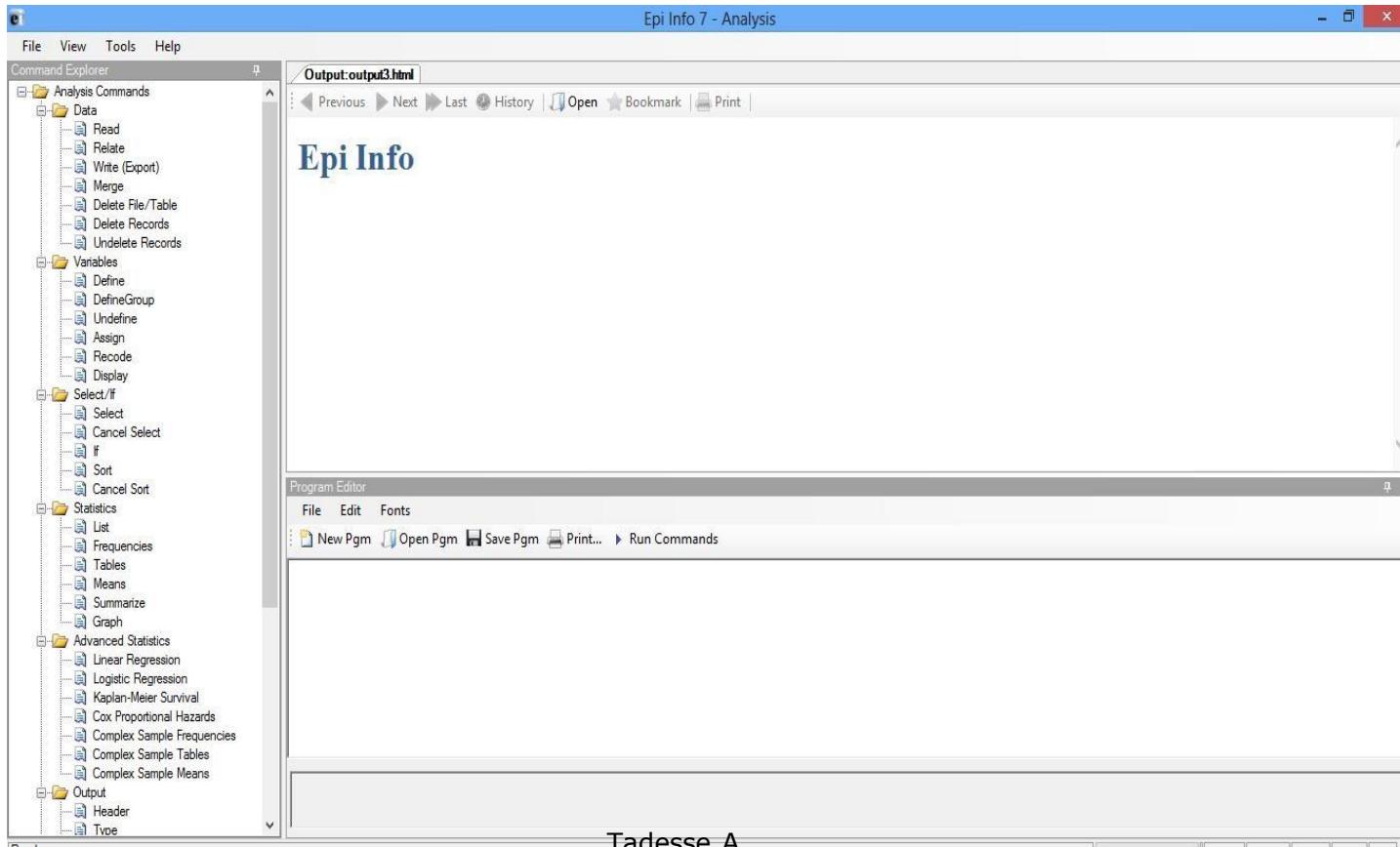
Tadesse A.

# Matched pair case control study



# Data analysis (1)

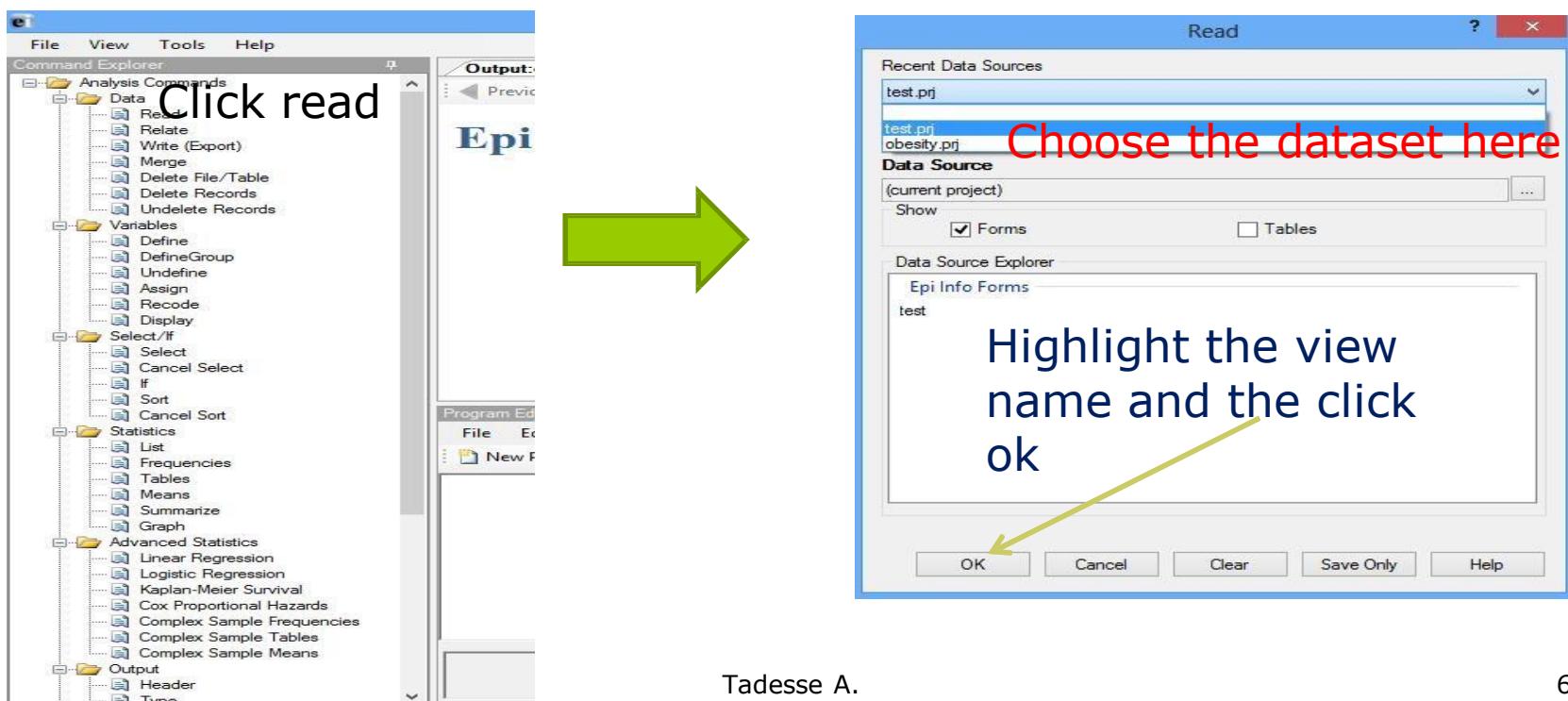
- Can be either classic or visual dashboard
  - Classic



Tadesse A.

## Data analysis (2)

- ❑ To start data analysis, first the data should be exported from access.
- ❑ Remember where you saved the data
- ❑ Click **classic** and you will see the following screen

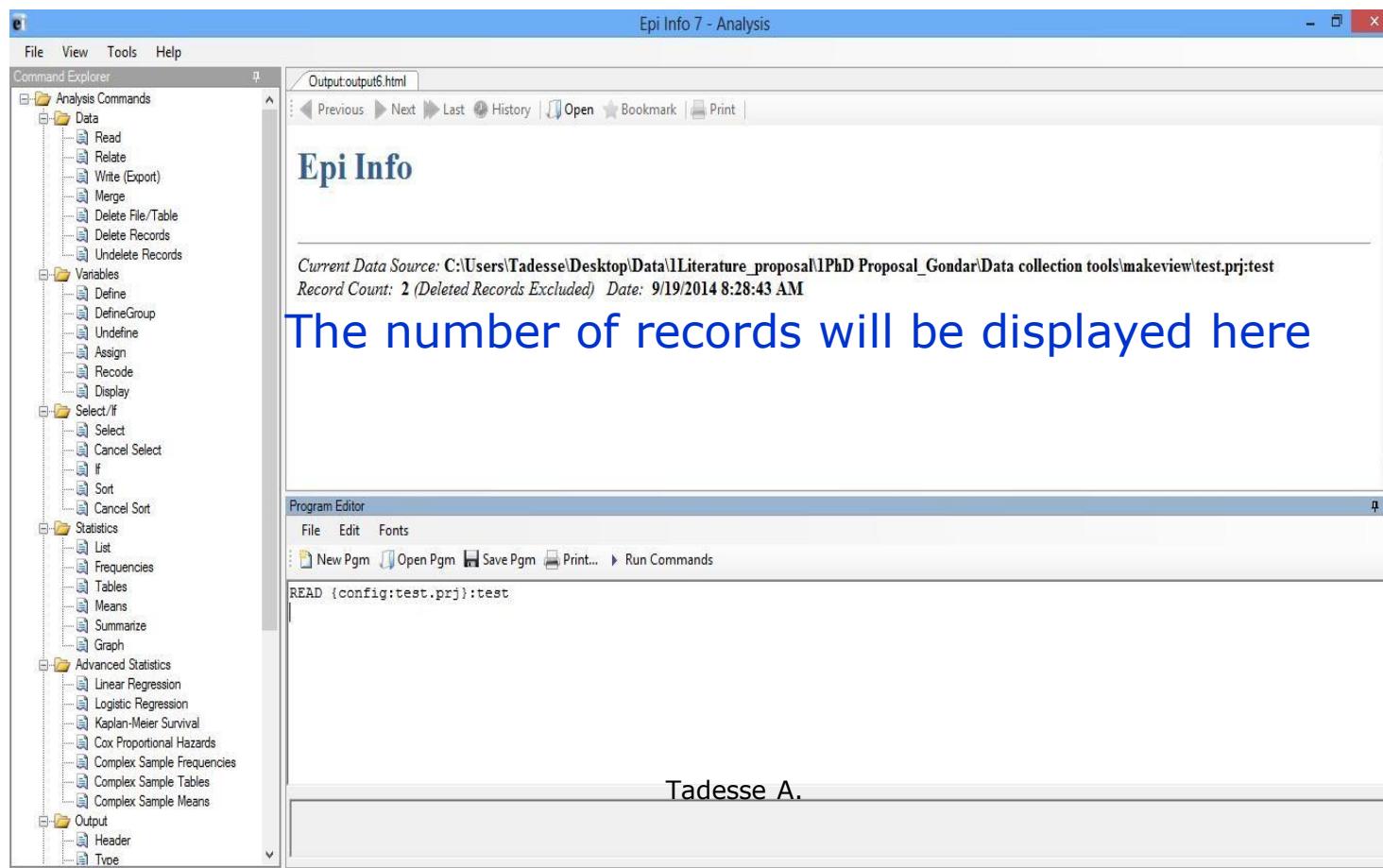


Tadesse A.

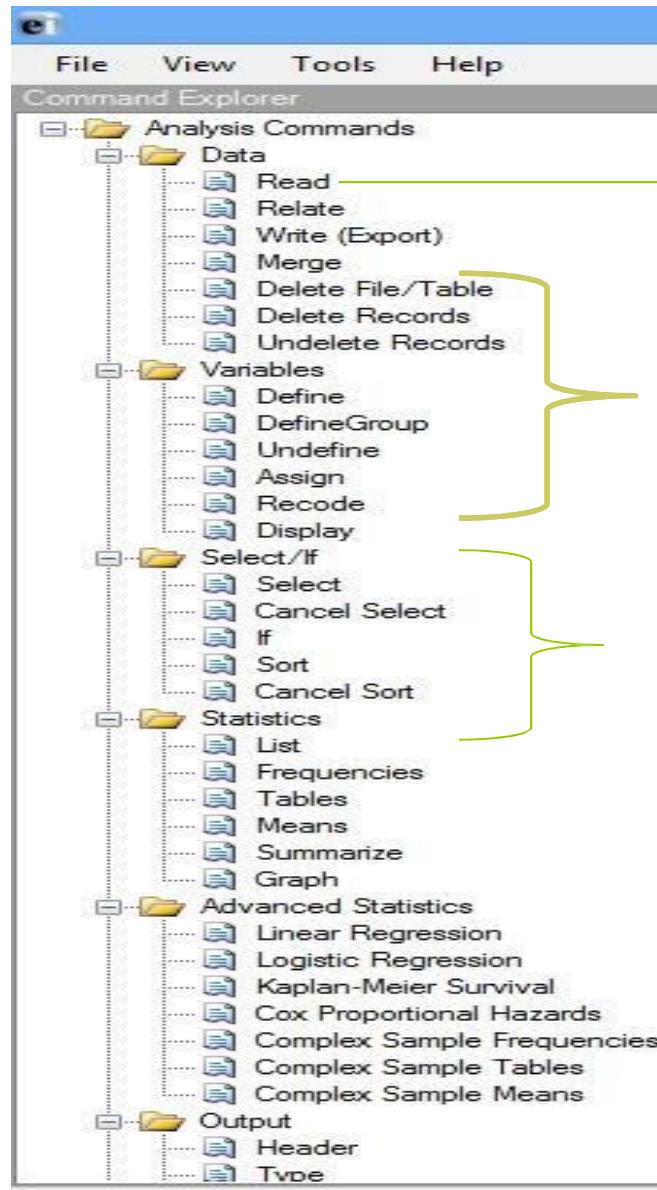
61

# Data analysis (3)

- After loading the data, we can start cleaning, coding and analysis (descriptive and analytic)



# Data analysis (4)



Import the data

Data management options

Descriptive analysis options

Advanced analysis options

# Exercises 3

---

- Read the data you entered
- Explore it
- Compute descriptive statistics
  - Frequency
  - Mean/s.d
  - Graphs

# Thank you!



# Session VII

# Introduction to SPSS



Tadesse Awoke (PhD)

Aug 2018

# Outline

---

- Introduction
- Installation
- Variable definition
- Data entry
- Data management
- Descriptive data analysis
- Inferential data analysis

# Introduction

---

- SPSS stands for **Statistical Package for Social Science**
- SPSS was made to be easier to use than other statistical software like STATA, S-Plus, R, or SAS.
- The newest version of SPSS is SPSS 22.0. Today we will be working on SPSS 20.
- SPSS for windows is a computer package that will perform a wide variety of statistical procedures.

# Introduction

---

- Using SPSS we can manipulate data, make graphs and perform statistical techniques varying from means to regression.
- Can be used for data entry, however it doesn't have any controlling mechanism for errors, rather it is pretty good for data analysis
- Advisable to use other software (EpiData or Epi-Info) for data entry and

# Installing SPSS v.20

---

- To work on SPSS first make sure you have install the software
- There are different version of SPSS starting from 10 to 22
- Installation is similar with other soft wares
- To install, get the soft ware in your PC
- Open the folder containing SPSS

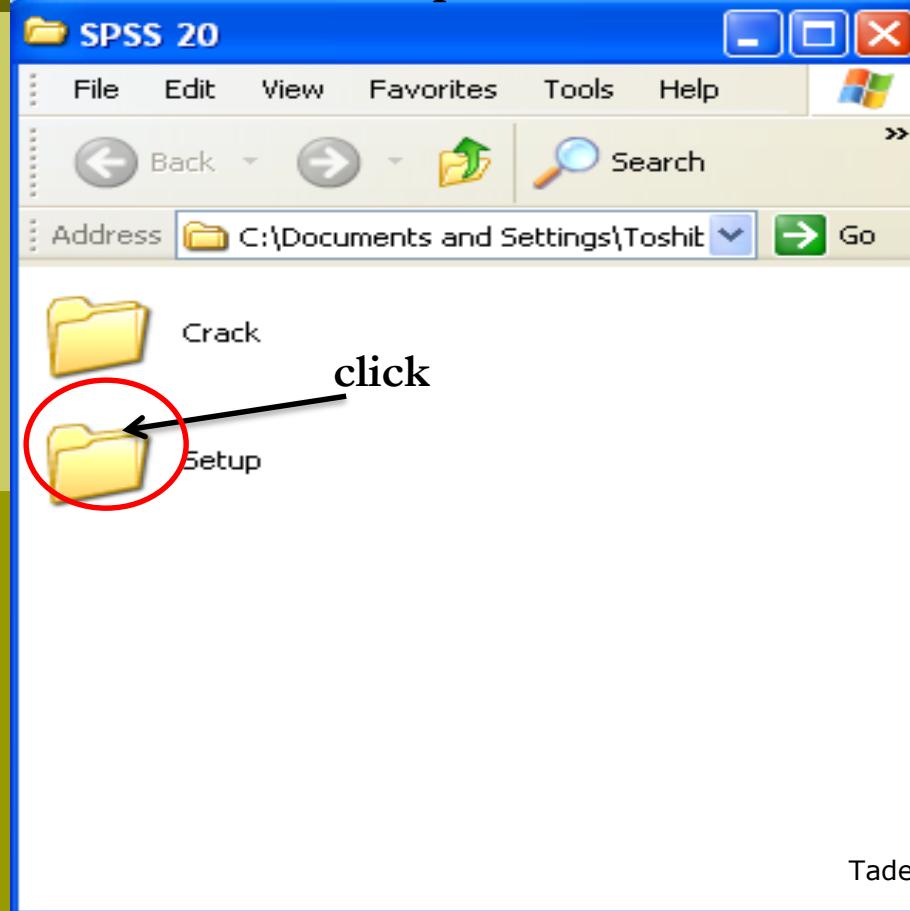


Tadesse A.

# Installing...

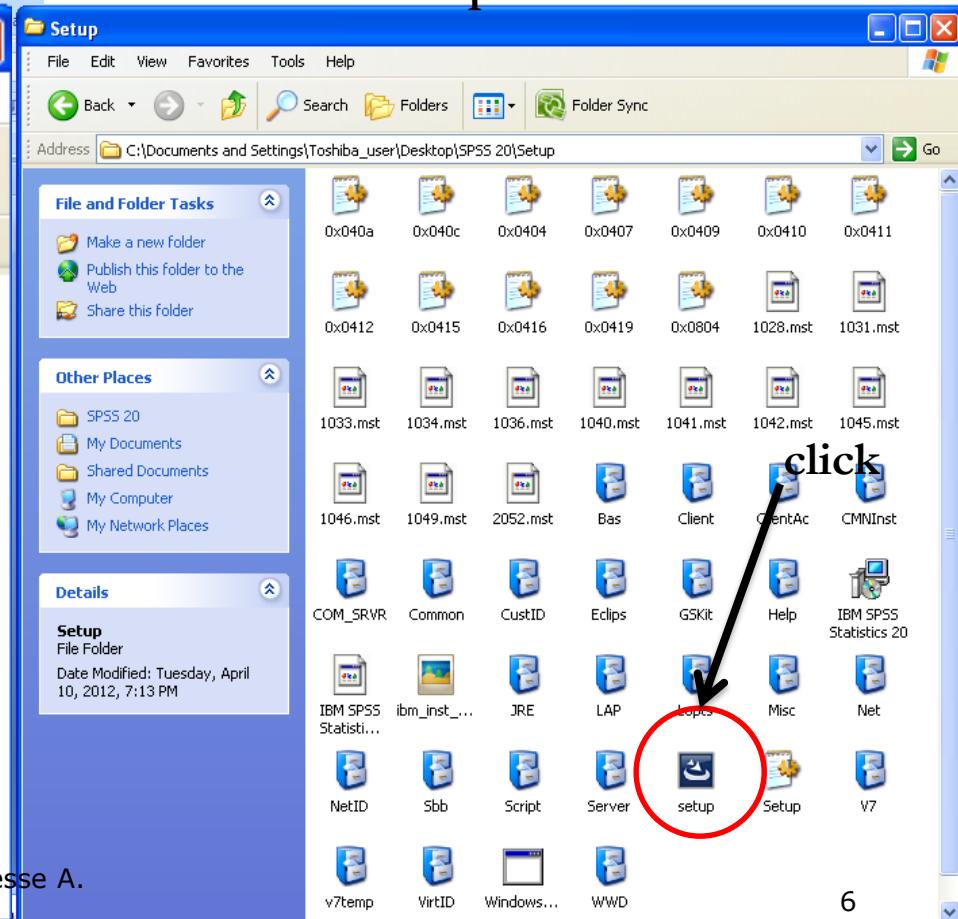
- After you open the folder, click setup

Step 1

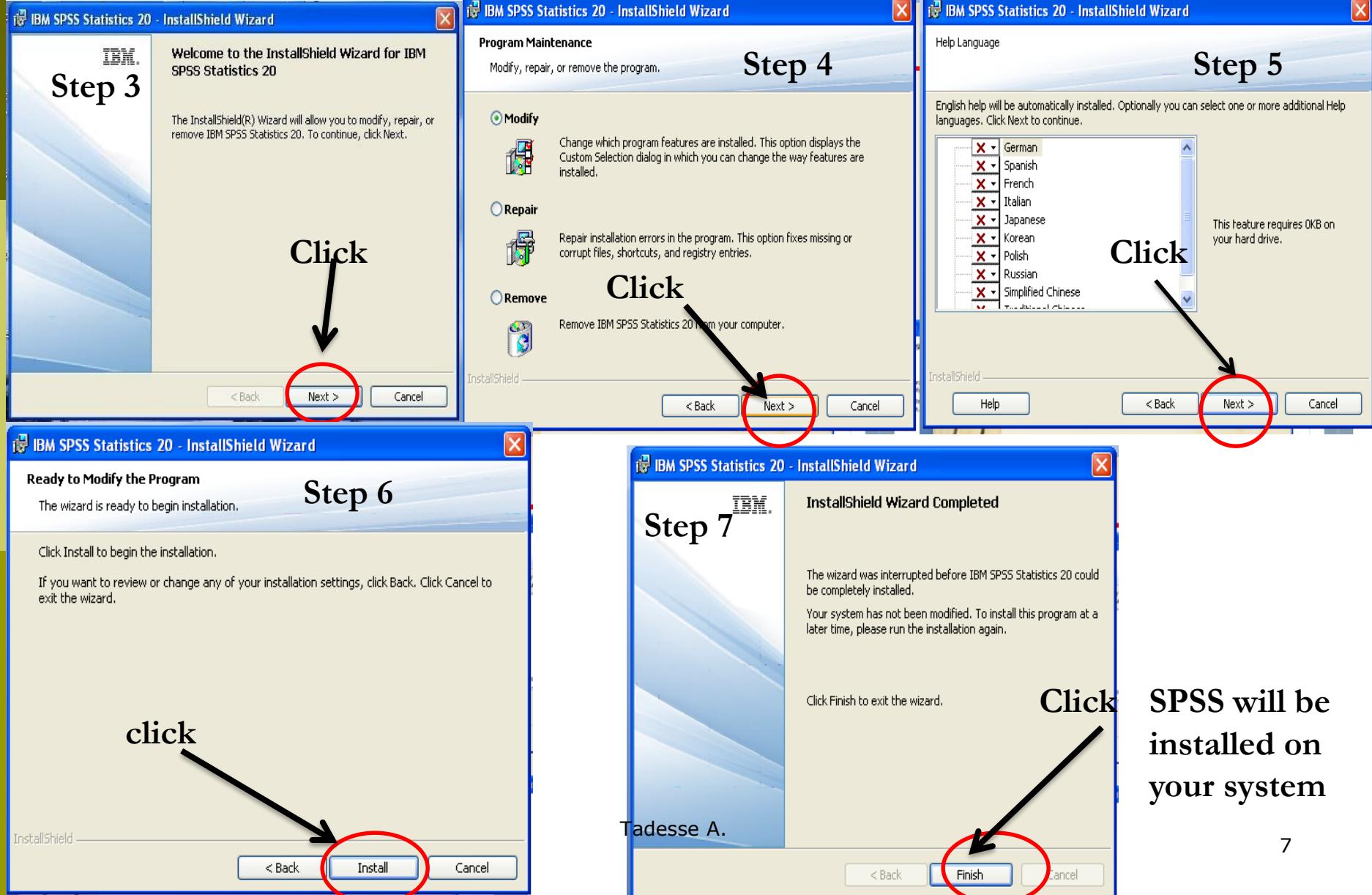


Tadesse A.

Step 2

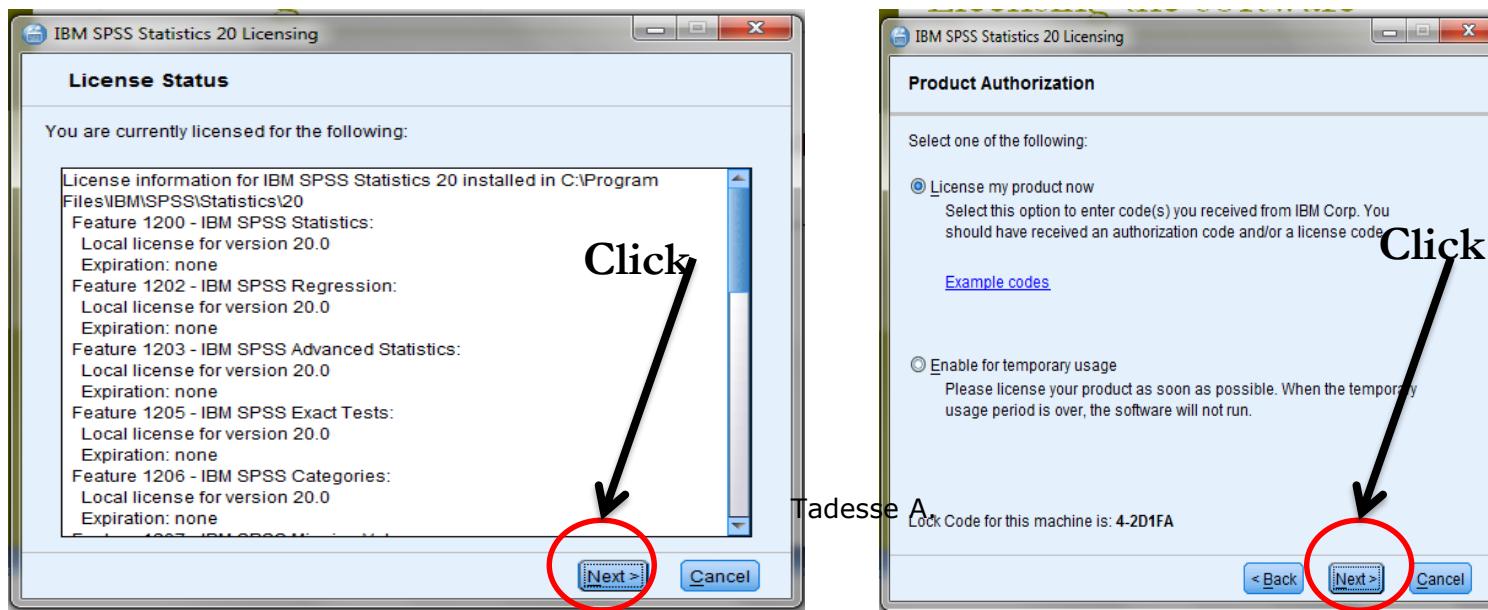


# Installing...



# Licensing the software

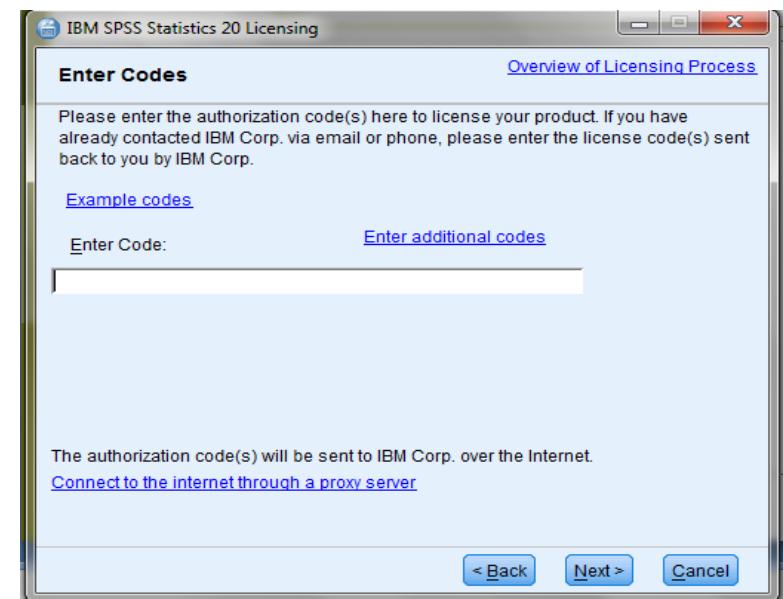
- ❑ Unless we activated the software by its authorization code, it will not work
- ❑ To make it work its magic, follow the following steps:
- ❑ Start → All Programs → IBM SPSS statistics→ IBM SPSS statistics 20 license Authorization



# Licensing the software

---

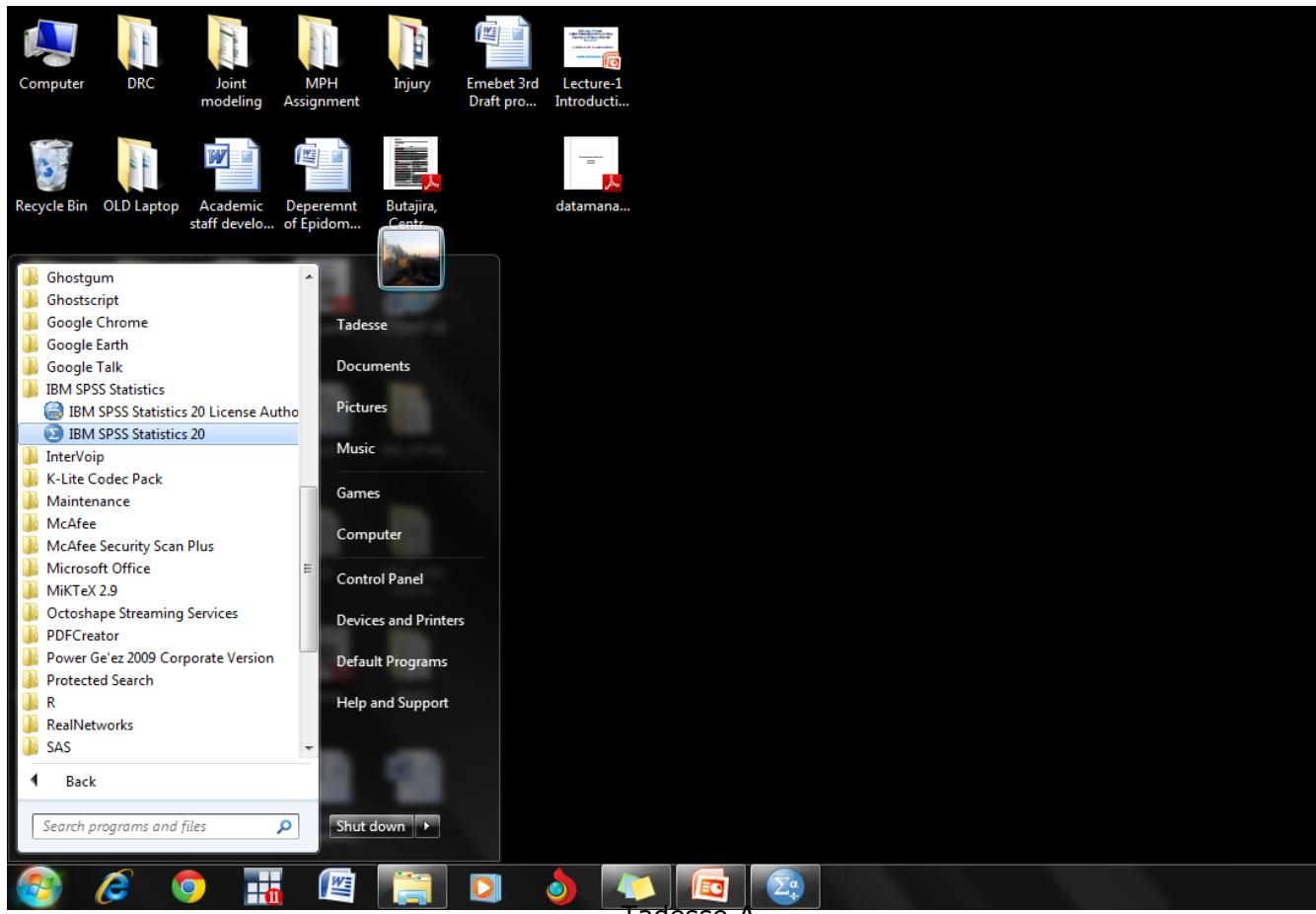
- ❑ Next go back to the SPSS folder and click “crack” and open “readme”
- ❑ Then copy the code next to “wing license code” using the License Authorization Wizard: and pest it in the “Enter code” space



- ❑ and click “Next”
- ❑ Now SPSS is ready to work!!!

# Steps to open SPSS

- Start → All Programs → IBM SPSS statistics → IBM SPSS statistics 20



Tadesse A.

# Types of SPSS windows

---

- Data editor
  - Variable view
  - data view
- Output viewer
- Syntax editor
- Script window

# Steps of Data analysis

What is the study design?

The research question

What is the outcome?

Are any groups being compared?

What are the variables of interest?

- Predictors or confounders?

## Step 5: Analyze

## Step 1: Set up

- Enter
- Import
  - Database
  - Excel, Access, Text

## Step 2: Inspect

- Defining variable properties
- Multiple response
- Utilities
  - Variable information

## Step 3: Clean

- Data preparation
  - Validation
  - Restructure data
- Transform (recoding)
  - Create new variables

Summarizing  
Tables  
Graphs

## Step 4: Describe

# Types of Data Base

---

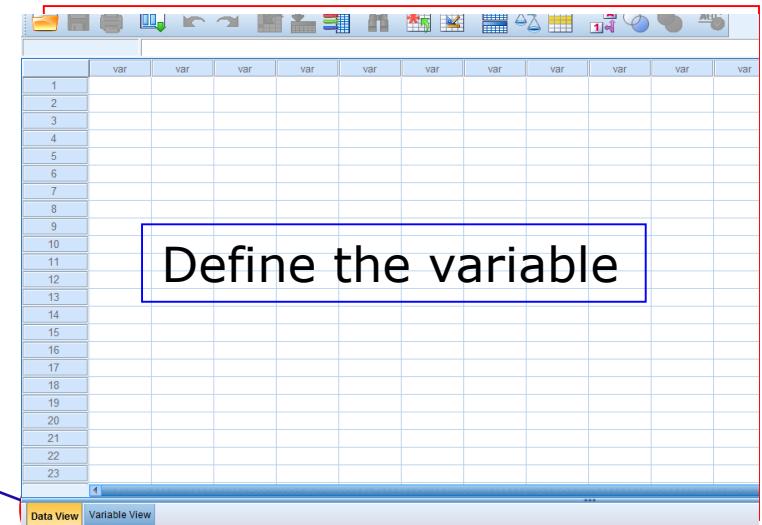
- Spreadsheets (e.g., Excel, SPSS Data Editor)
  - Prone to error, data corruption, & mismanagement
  - Lack data controls, limited programmability
  - Suitable only for small and didactic projects
  - Also good for last step data cleaning
- Public domain programs (e.g., EpiData, Epi Info)
  - Controlled data entry, good programmability
  - Suitable for research and field use

# SPSS Program Windows

## □ SPSS Program Windows

### ■ Data Editor

#### □ Data View



#### □ Variable View

The screenshot shows the Variable View window of SPSS. It displays a table of variables with 24 rows, numbered 1 to 24. The columns represent various properties: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and More. A blue rectangular box is drawn around the 'Values' column, containing the text 'Data entered here'. A red circle highlights the last row of the table, which is labeled 'n/c'.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	More
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
n/c									

Tadesse A.

# SPSS Program Windows

## ■ Output Viewer

The screenshot shows the SPSS Output Viewer window. On the left, the 'Output' tree view shows a node for 'Frequencies'. The main pane displays two tables: one for 'sex' and one for 'have you ever had an abortion?'. A red diagonal arrow points from the text 'Output displayed' towards the tables.

**Statistics**

	sex	have you ever had an abortion?	total live birth
N	8050	8040	8040
Valid		0	10
Missing			10

**Frequency Table**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid .00	3944	49.0	49.0	49.0
1.00	4106	51.0	51.0	100.0
Total	8050	100.0	100.0	

**have you ever had an abortion?**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.00	616	7.7	7.7	7.7
2.00	7424	92.2	92.3	100.0
Total	8040	99.9	100.0	
Missing System	10	.1		
Total	8050	100.0		

## ■ Syntax Editor

The screenshot shows the SPSS Syntax Editor window. The left pane shows the code history with entries for activating and closing datasets. The right pane contains a large red box labeled 'Code used to compute'.

```
FILE HANDLE Data1 FILENAME='C:\Users\Tadesse\Desktop\Training_Materials\So'.
DATASET ACTIVATE DataSet1.
DATASET CLOSE DataSet2.
```

Tadesse A.

# Variable view



## Defining a Variable

---

- Variables are defined in the **Variable View** within the **Data Editor**
- Within **Variable View**, rows correspond to the variables in the data file, and columns correspond to their defining characteristics.
- Selecting a cell allows the corresponding characteristic to be specified or changed, either by over-typing or via a scrollable list.

# Variable view

- The default window will have the data editor
- There are two sheets in the window:

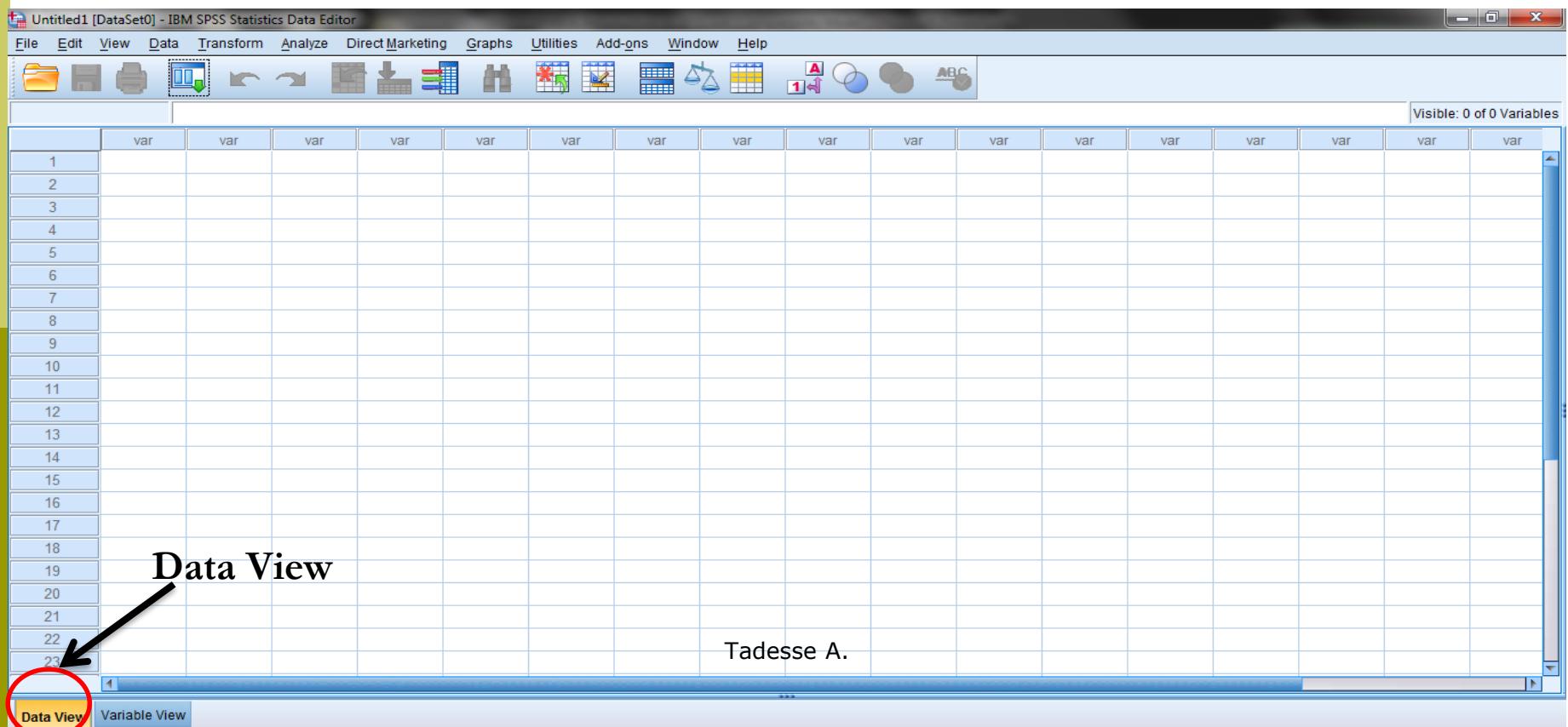
1. Variable view

2. Data view

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads "Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. Below the menu is a toolbar with various icons. The main area is a grid for defining variables, with rows numbered 1 to 24. A red box highlights the first row, which contains columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. An arrow points from the text "Variable view" to the top-left corner of this highlighted row. At the bottom of the screen, a tab bar shows "Data View" and "Variable View", with "Variable View" highlighted by a red circle.

# Data View

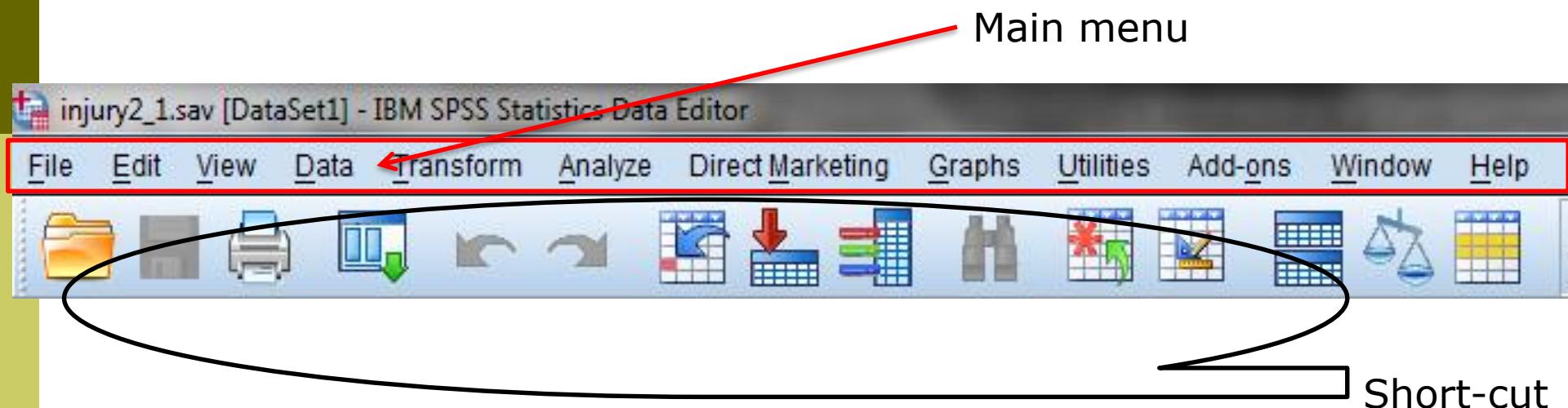
- This sheet is visible when you first open the Data Editor and this sheet contains the data
- Click on the tab labeled data View



# The basics of managing data files



# The main menu bar



**File:** These are the basic file management operations. e.g. opening, saving, and printing files

# Edit

---

- This allows you to perform editing functions on the current data set.
- e.g. cut, copy, clear, undo changes and redo changes

The screenshot shows the IBM SPSS Statistics interface with the 'injury2\_1.sav [DataSet1] - IBM SPSS Statistics' window title. The 'Edit' menu is open, displaying various editing functions. The 'Insert Variable' option is highlighted with a yellow background. Below the menu, the 'Data View' table is visible, showing data for variables like Age, Sex, and Marital Status for cases 1 through 22. The 'Variable View' tab is also visible at the bottom.

Case	Age	Sex	Marital Status
1	20.00	1.	Male
15	20.00	1.	Male
16	40.00	1.	Male
17	32.00	1.	Male
18	48.00	1.	Male
19	37.00	1.	Male
20	20.00	1.	Male
21	30.00	1.	Male
22	45.00	1.	Male
--			

Data View      Variable View

# View

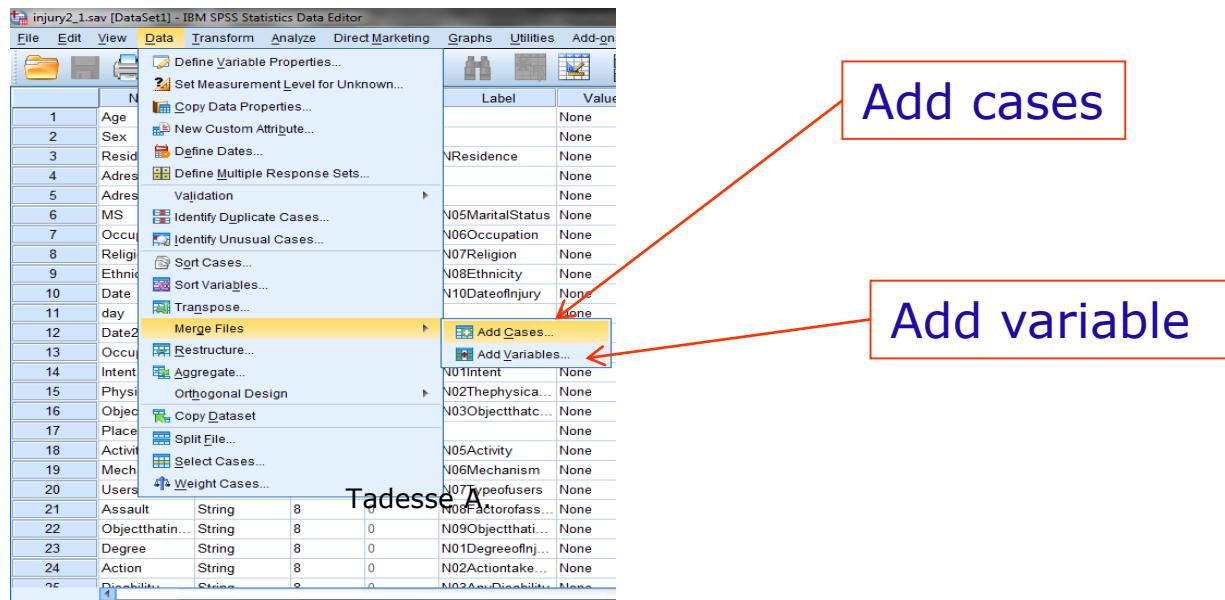
- Allows you to change the current view of data, as well as toolbar options; e.g. grid lines, value labels

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads "injury2\_1.sav [DataSet1] - IBM SPSS Statistics Data Editor". The menu bar is visible with "File", "Edit", "View" (which is highlighted in yellow), "Data", "Transform", "Analyze", and "Direct". A sub-menu for "View" is open, listing several options: "Status Bar", "Toolbars", "Menu Editor...", "Fonts...", "Grid Lines", "Value Labels" (which is selected and highlighted in yellow), "Mark Imputed Data", and "Customize Variable View...". Below the menu, there is a data grid with 22 rows and 4 columns. The first column contains row numbers 1 through 22. The second column contains values 50.00, 27.00, 45.00, 40.00, 35.00, 25.00, 23.00, 30.00, 20.00, 40.00, 32.00, 48.00, 37.00, 20.00, 30.00, and 45.00. The third and fourth columns both contain two entries: "1. Male" and "2. Rural". At the bottom left, the text "Tadesse A." is visible. At the bottom right, there are buttons for "Data View" (highlighted in yellow) and "Variable View".

1	50.00	1. Male	2. Rural
2	27.00	1. Male	2. Rural
3	45.00	2. Femal	1. Urban
4	40.00	1. Male	1. Urban
5	35.00	2. Femal	2. Rural
6	25.00	1. Male	2. Rural
7	23.00	1. Male	1. Urban
8	30.00	1. Male	1. Urban
9	20.00	1. Male	1. Urban
10	40.00	1. Male	2. Rural
11	32.00	1. Male	2. Rural
12	48.00	1. Male	2. Rural
13	37.00	1. Male	2. Rural
14	20.00	1. Male	2. Rural
15	30.00	1. Male	2. Rural
16	45.00	1. Male	2. Rural
17	32.00	1. Male	2. Rural
18	48.00	1. Male	2. Rural
19	37.00	1. Male	2. Rural
20	20.00	1. Male	2. Rural
21	30.00	1. Male	2. Rural
22	45.00	1. Male	2. Rural

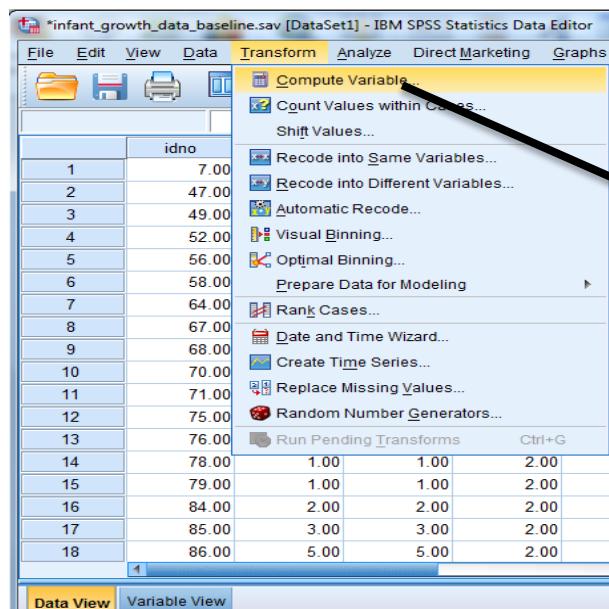
# Data: Merge files

- In some cases, a survey questionnaire may comprise many variables.
- We may divided the questionnaire and use different clerks to enter the data.
- Finally, we can merge to make the data ready for analysis using “Add variable” procedure



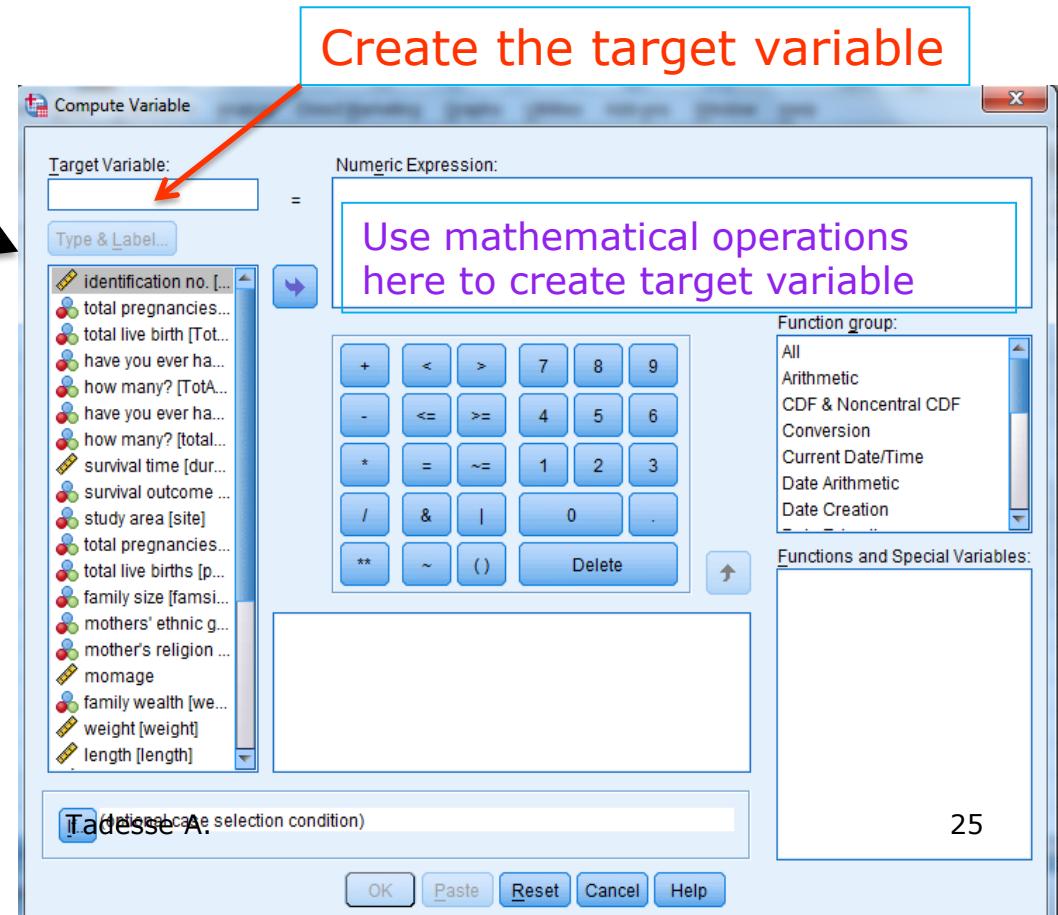
# Transform: compute variable...

- Variable can be computed using the four mathematical operations (+, -, × or ÷)



The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu bar is visible at the top, and the 'Transform' menu is highlighted. A black arrow points from the 'Compute Variable' option in the 'Transform' menu down to the 'Compute Variable' dialog box. The main window displays a data view with several columns and rows of numerical data.

	idno		
1	7.00		
2	47.00		
3	49.00		
4	52.00		
5	56.00		
6	58.00		
7	64.00		
8	67.00		
9	68.00		
10	70.00		
11	71.00		
12	75.00		
13	76.00		
14	78.00	1.00	1.00
15	79.00	1.00	1.00
16	84.00	2.00	2.00
17	85.00	3.00	3.00
18	86.00	5.00	5.00



## Transform: compute variable...

---

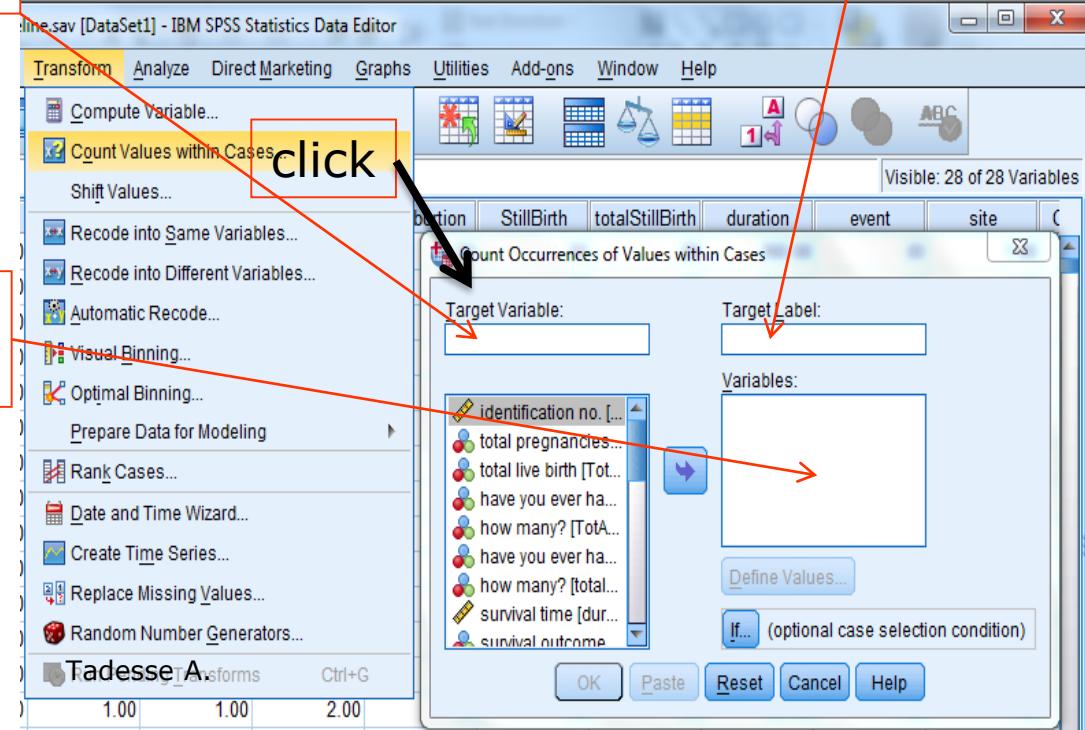
- Example:
- 
- We can compute BMI from height and weight
  - $BMI = \text{weight}/(\text{height} * \text{height})$
- Logarithm transformed CD4 count
  - $\ln CD4 = \log(CD4)$
- And so on

# Transform: Count values with cases...

- The variables with a certain value can be counted and new variable can be generated

Defined the target variable here

State target level here



Insert the variables in which the value will be counted

## Transform: Count values with cases...

---

- It is useful when we want to compute knowledge, practice, attitude, etc by the list of questions
- After we compute the target variable, we can set criteria to create the final variable for analysis
  - E.g above the mean (median or %) --- knowledgeable
  - below the mean (median or %)---- not knowledgeable

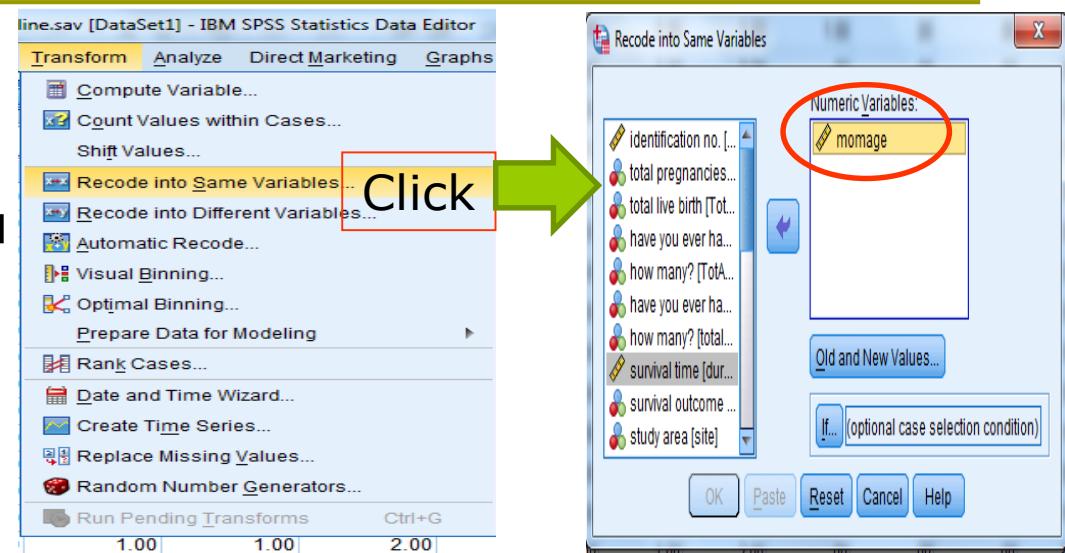
## Transform: Recoding Variables

---

- Recoding, rennumbers or collapses the values of a variable.
- Two options
  - Recode into different variables
  - Recode into different variables

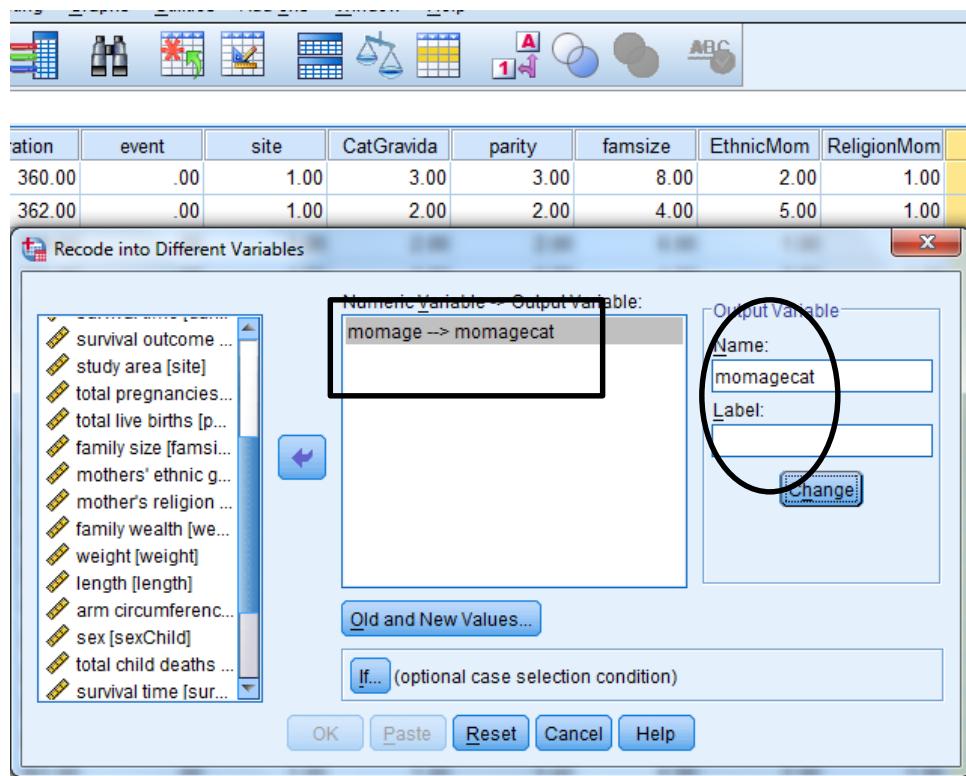
# Recode in to the same variable

- Click recode in to same variable
- Highlight the variable you want to recode and move over with arrow
- Click **Old and New Values**



# Recode in to the different variable

- This procedure has similar steps with “recode in to the same variable”.
- The basic difference here is, new variable will be generated instead of changing the existing variable
- Thus, we need to give name and label for the new variable and click “change”
- Eg. MomageCat for the variable momage



# Data analysis



1. Descriptive data analysis
2. Inferential data analysis

# Analyze: Descriptive Data Analysis

- Assuming that the data is ready for analysis, we start exploring using descriptive analysis

The screenshot shows the SPSS software interface. A black callout arrow points from the text "Analyze ---->" to the "Analyze" menu item in the top menu bar. The "Analyze" menu is open, displaying a list of statistical analysis options. The menu bar also includes File, Edit, View, Data, Transform, Direct Marketing, and Graphs. The main window displays a data view with two columns: idno and TotPreg, containing 20 rows of data. The bottom right corner of the slide has the number 33.

idno	TotPreg
1	7.00
2	47.00
3	49.00
4	52.00
5	56.00
6	58.00
7	64.00
8	67.00
9	68.00
10	70.00
11	71.00
12	75.00
13	76.00
14	78.00
15	79.00
16	84.00
17	85.00
18	86.00
19	91.00
20	92.00

Tadesse A. 33

# Inferential data analysis (1)

---

- We have seen different procedure to explore the data descriptively
- Descriptive output only tell us the characteristics of the sample data
  - Don't tell us the characteristics of the population
  - If the objective of the analysis is to say something about the population characteristics, inferential analysis is needed

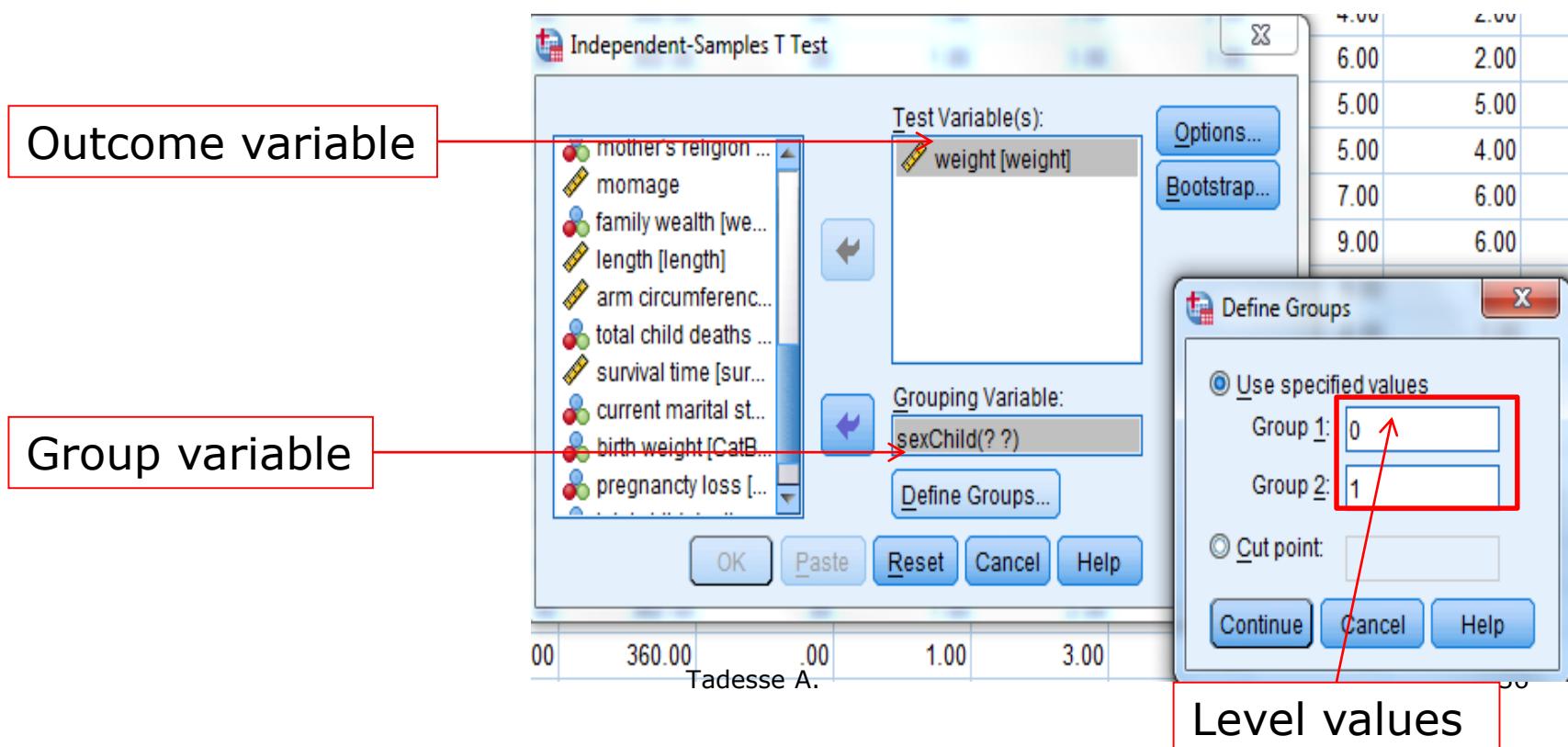
# Inferential data analysis (2)

---

- These includes
- Comparison of means
  - t-test
    - One sample t-test
    - Independent sample t-test
    - Paired sample t-test
  - ANOVA
- Regression
  - Linear regression
    - LMM
    - Poisson regression
  - Logistic regression
    - GEE, GLMM, ...
- Survival analysis, ...

# Independent sample t-test

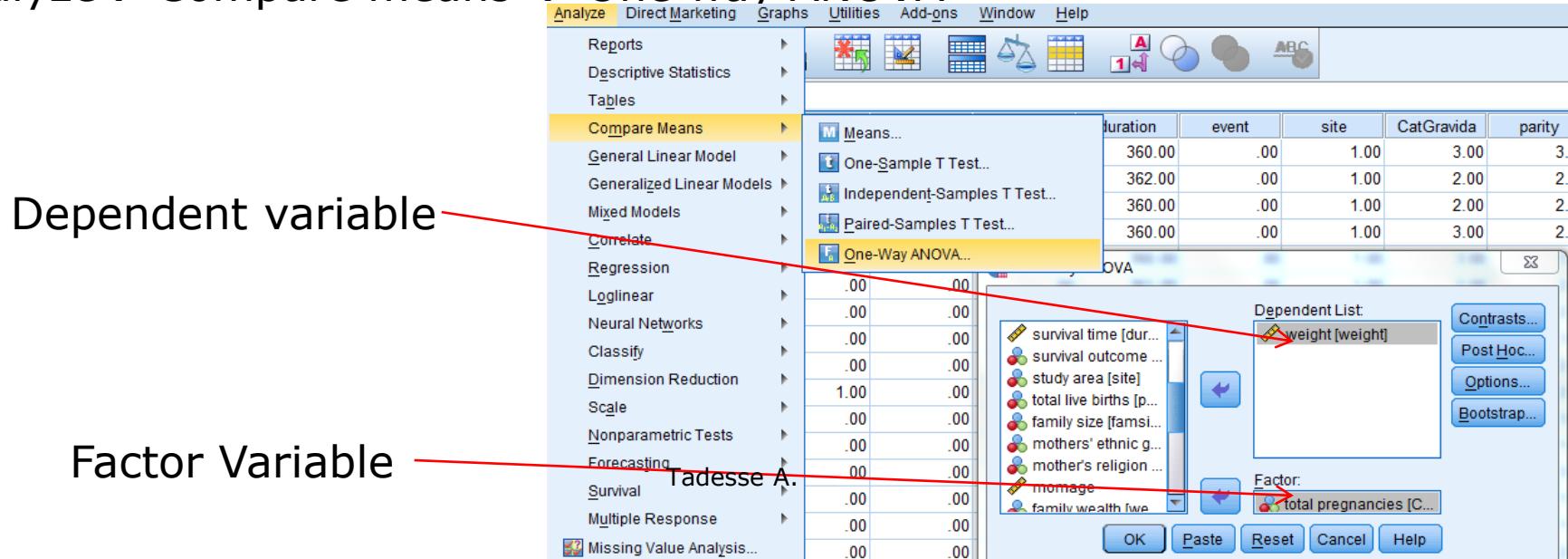
- To compare the outcome variable among two levels of factor variable
  - Eg. Comparing birth weight among sex of the child



# Analysis of Variance (ANOVA)

- ❑ If the factor/group variable has more than two levels, t-test is not useful to compare the outcome variable between the groups
- ❑ Analysis of Variance (ANOVA) is used to compare between groups if more than two

Analyze → Compare means → One way ANOVA

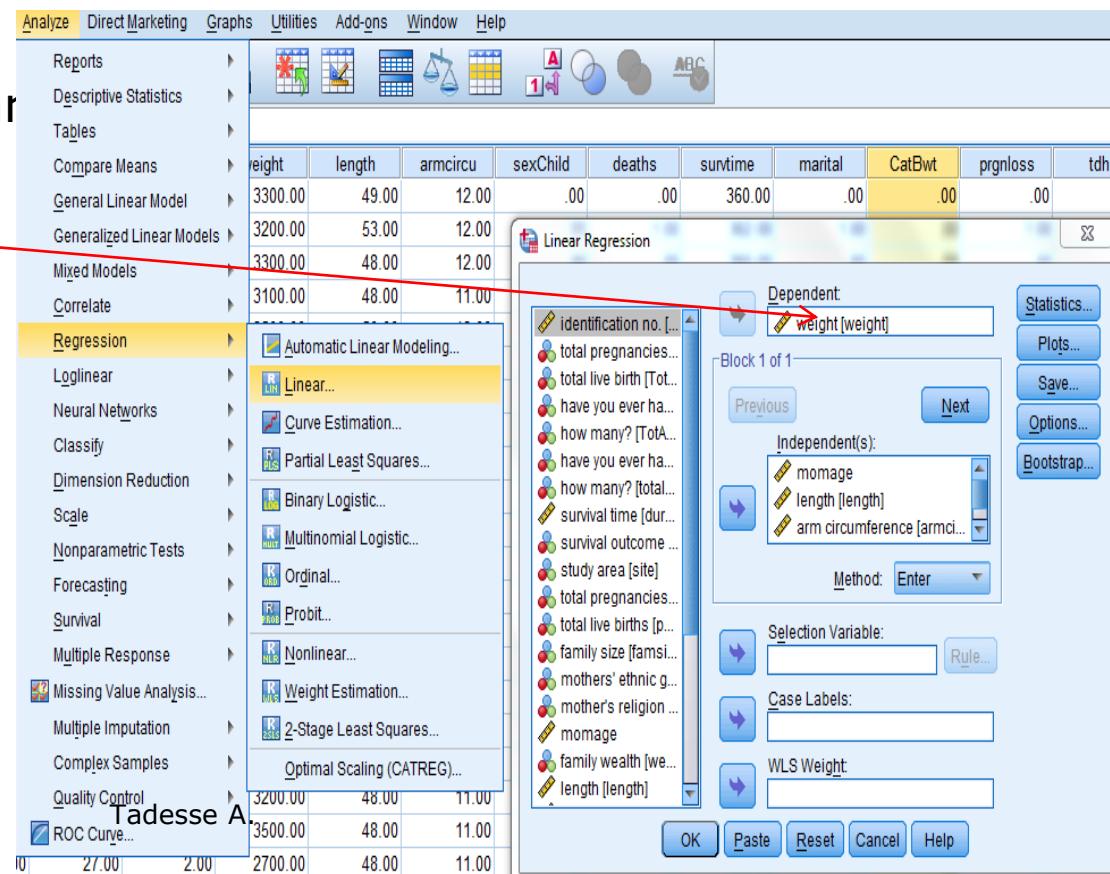


# Linear regression

- We may be interested in predicting the effect of independent variables on the outcome variable which is continuous

Analysis → regression → linear

Dependent variable



# Logistic regression

---

- Now, the outcome variable is categorical
- If the categorical outcome variable is binary (only 2 possible values), then binary logistic regression is fitted
- If the outcome variable has more than two levels of nominal, then multinomial logistic regression
- If the outcome variable has ordered levels, then ordinal logistic regression is fitted

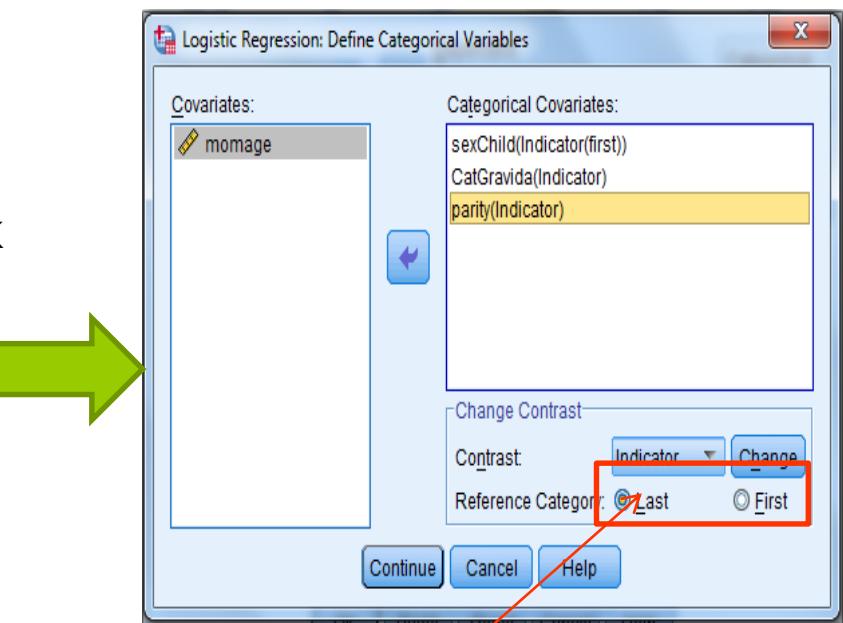
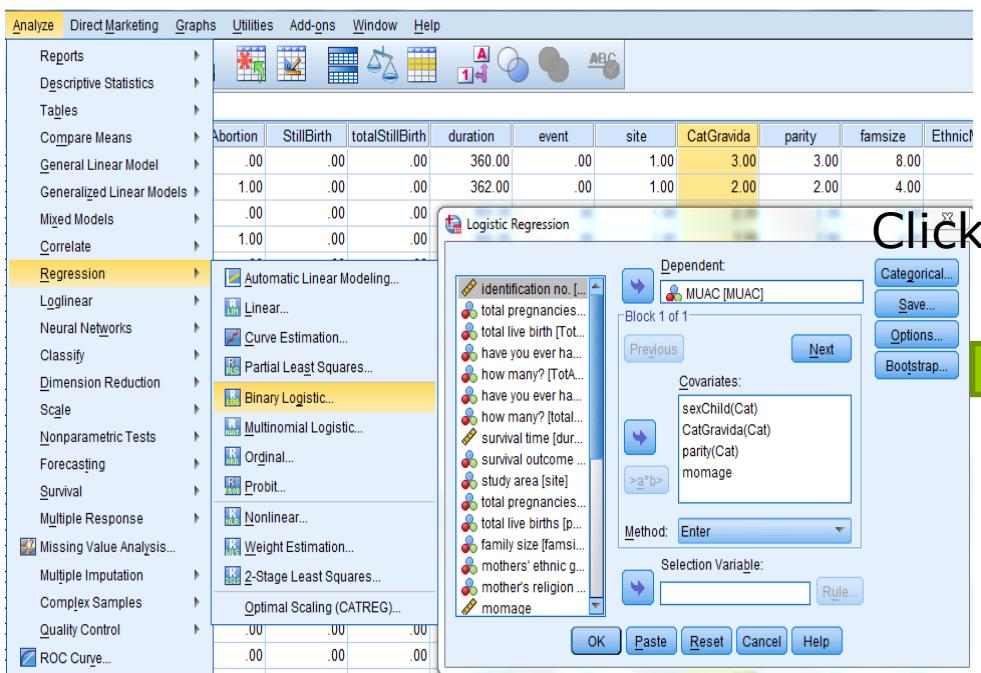
# Binary logistic regression

---

- **Dependent variable:** Birth weight (MUAC, normal=1, underweight =0)
- **Explanatory variable:** Age of mother, sex of the child, gravidity, parity, marital status, and so on
  - **Univariate logistic regression:** Fitting logistic regression for each explanatory variable
  - 
  - Odds ratio, p5% CI and p-value will be computed
  - A variable will be taken to multivariable logistic regression if p-value is less than 20%

# Binary logistic regression

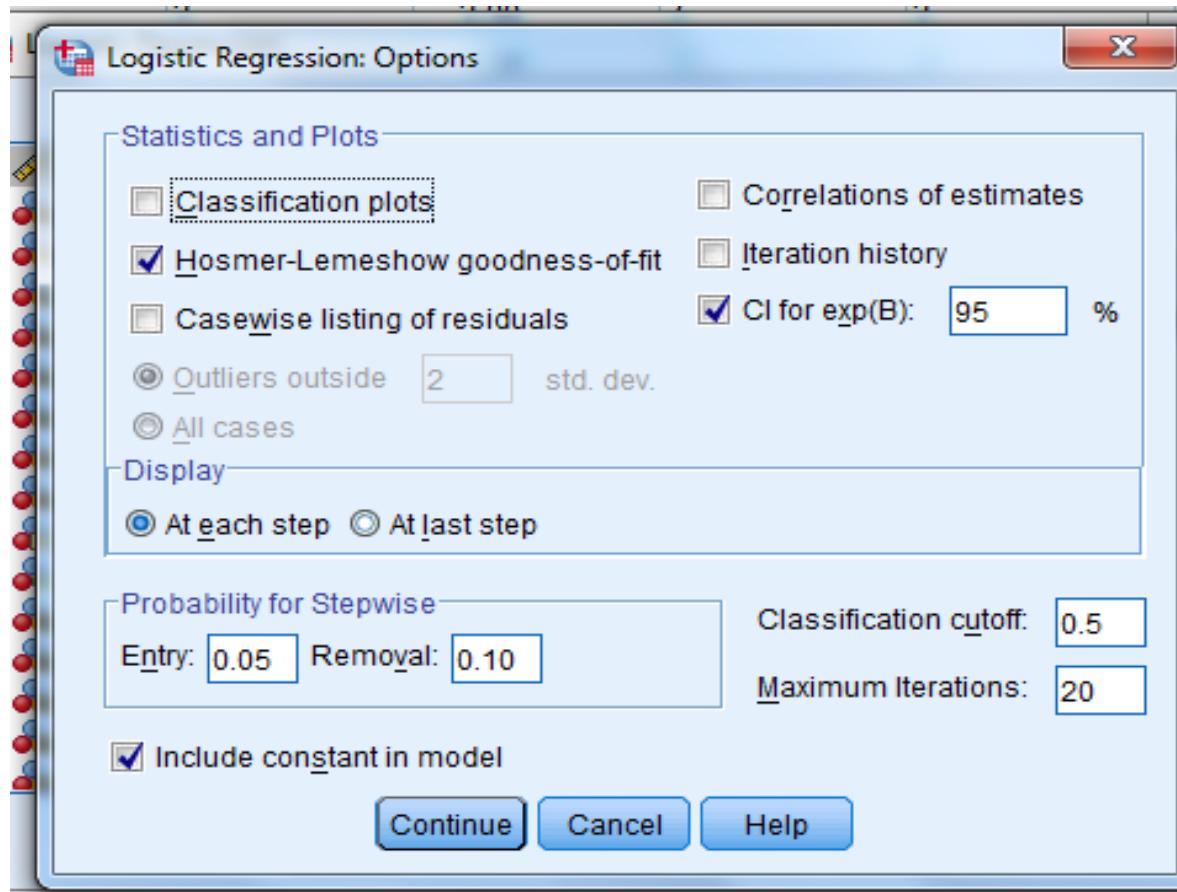
- Reference level should be defined by clicking “**categorical**” and choose either last or first



Tadesse A.

We can change the reference level by choosing either last or first, then click "change" <sup>41</sup>

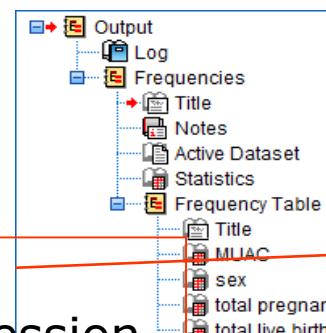
# Binary logistic regression



# Univariate logistic regression

- First run the frequency and check the distribution of each explanatory variable over the dependent variable

This tells us whether the Variable is eligible for regression



Frequency Table

MUAC				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid <11.49	6016	74.7	77.8	
>11.5	1721	21.4	22.2	
Total	7737	96.1	100.0	
Missing System	313	3.9		
Total	8050	100.0		

sex

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid .00	3944	49.0	49.0	49.0
1.00	4106	51.0	51.0	100.0
Total	8050	100.0	100.0	

total pregnancies

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	1804	22.4	22.4	22.4
2-4	3585	44.5	44.6	67.0
>4	2651	32.9	33.0	100.0
Total	8040	99.9	100.0	
Missing System	10	.1		
Total	8050	100.0		

Tadesse A.

# Univariate logistic regression

	Variables in the Equation						P-value	95% C.I. for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper	
Step 1 <sup>a</sup>	.219	.055	15.970	1	.000	1.245	1.118	1.387	
Constant	-1.367	.040	1147.459	1	.000	.255			

a. Variable(s) entered on step 1: sexChild.

- As we see from p-value (<0.05) or
- 95% CI (failed to consist 1), sex of the child has significant association with birth weight of the new born
- The Odds ratio here is **unadjusted**, which may be affected by confounding variables
- However, in the univariate analysis we only focus on 20% p-value

# Multivariable logistic regression

- Now we are going to include all potential explanatory variables in the model simultaneously

Explanatory variables with codes used for the analysis

Categorical Variables Codings

	Frequency	Parameter coding	
		(1)	(2)
total live births	1	1.000	.000
	2-4	.000	1.000
	>4	.000	.000
total pregnancies	1	1.000	.000
	2-4	.000	1.000
	>4	.000	.000
sex	.00	.000	
	1.00	1.000	

Hosmer and Lemeshow Test

→

Step	Chi-square	df	Sig.
1	7.963	8	.437

Goodness of fit test of the model can be checked using Hosmer and Lemeshow test (higher p-value---good model)

Tadesse A.

# Multivariable logistic regression

- Here, the cutoff value is set at 0.05 p-value

The resulting odds ratio is **adjusted** and less affected by confounding variable

	Variables in the Equation					P-value	Adjusted odds ratio	
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
Step 1 <sup>a</sup>	.224	.055	16.433	1	.000	1.251	1.122	1.393
sexChild(1)			2.581	2	.275			
CatGravida			2.485	1	.115	.577	.291	1.143
CatGravida(1)	-.550	.349						
CatGravida(2)	-.139	.205	.463	1	.496	.870	.582	1.300
parity			.196	2	.907			
parity(1)	-.078	.348	.050	1	.823	.925	.468	1.829
parity(2)	.035	.207	.029	1	.865	1.036	.691	1.554
momage	.001	.006	.024	1	.876	1.001	.989	1.013
Constant	-1.223	.209	34.152	1	.000	.294		

a. Variable(s) entered on step 1: sexChild, CatGravida, parity, momage.

# The End

