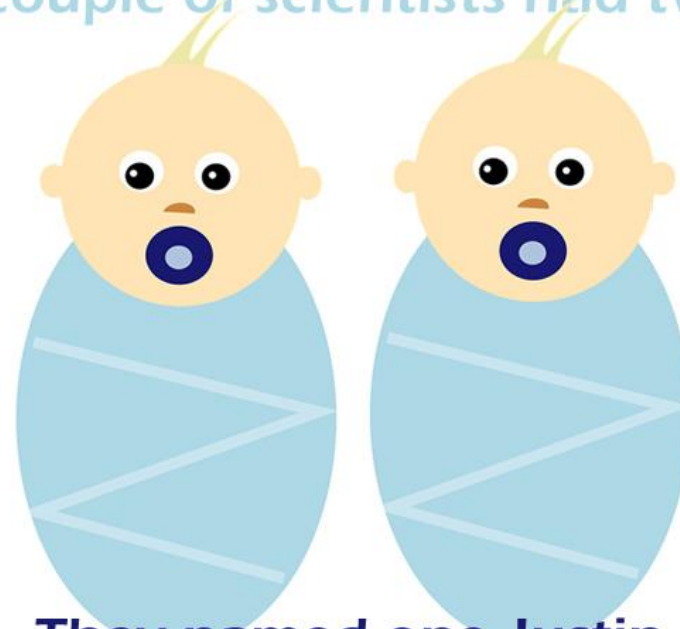


A couple of scientists had twins



They named one Justin
and the other Control

Author: Conf. Dr. Cosmina-Ioana Bondor

Lecture 3 – Study protocol, Case-control study



ALWAYS

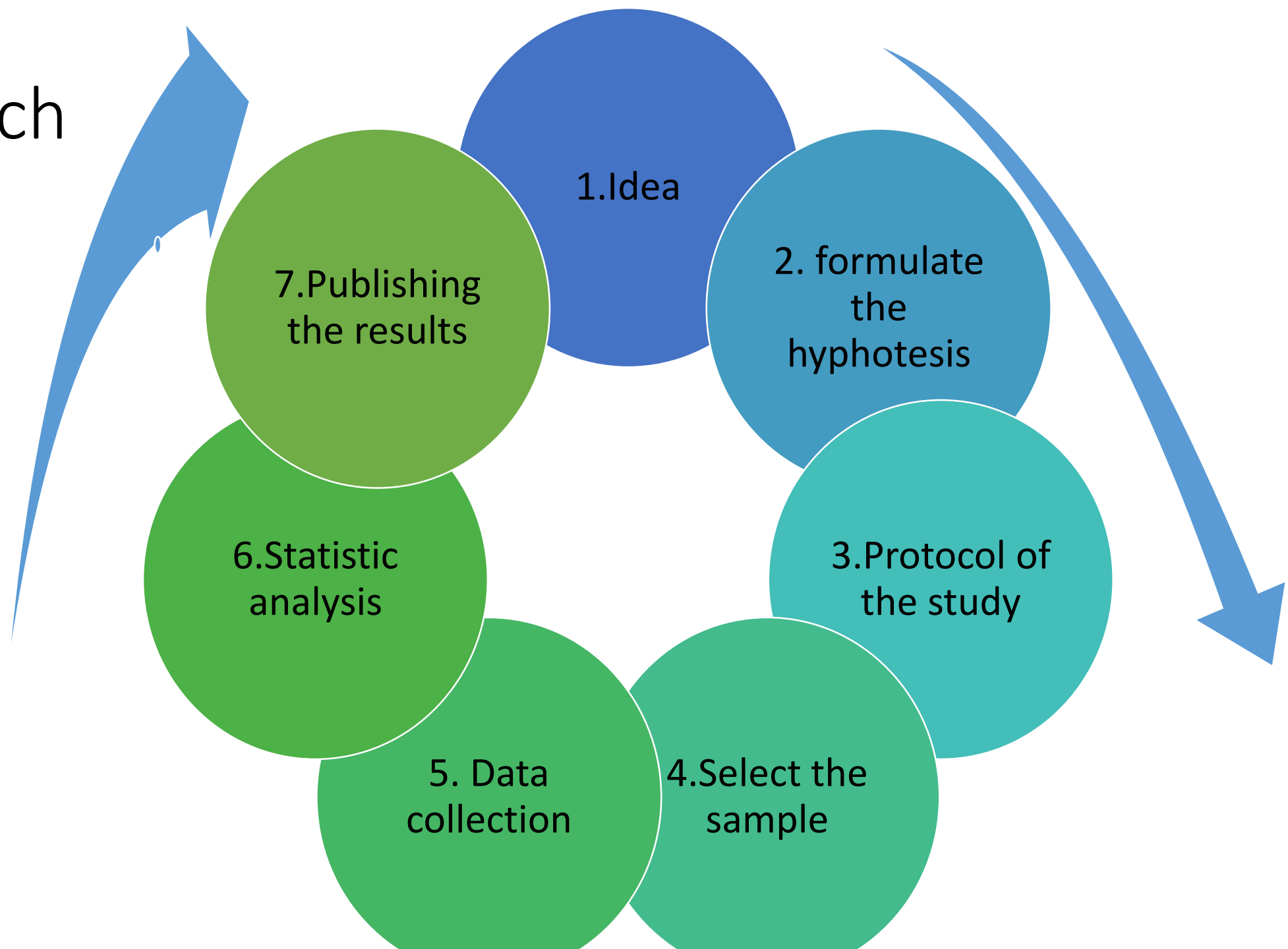


SEEK



KNOWLEDGE

Research



Hypothesis = purpose (aim) of the
study

Aim of the study

- a more general aspect

Examples:

- Evaluating an association between
 - a risk factor and a disease
 - a treatment and an outcome
 - between two factors
- Comparing
 - two diagnostic techniques
 - two treatments

Objective – what we want to demonstrate

Main objective

- Study hypothesis – what will we study?
- Ex. Assess the association between alcohol consumption during pregnancy and the occurrence of a malformation in the child

Secondary objectives (what else can we study?)

- Ex. Quantify the importance of this link



Study objectives

= precise, practical steps

Major objectives:

- existence of a link between the risk/prognostic factor and the disease
- quantification of the importance of the link between the risk/prognostic factor and the disease
- difference between treatments
- difference between diagnostic tests
- quantification of the difference

Secondary objectives of the study.

- Other biological phenomena studied in the same study

3. Protocol – plan of the research

Study protocol

a document

- includes the study description
 - motivation (why are we doing the study?)
 - purpose
 - objectives
- patients/subjects
 - description of participating groups
- methods
 - how do we diagnose?
 - how do we measure?

Study protocol

- Where will it take place?
- How will it take place?
- Who will be included/excluded in the study?
- What will we measure/investigate/observe?
- What measurement/investigation methods will be used?
- What devices will we investigate/measure/observe with?
- What will be the data analysis methods?
- Who will be the investigators?
- How will they be trained?
- etc.
- The protocol answer to all this questions

Study protocol

- The answer to all these questions
 - but also
 - the motivation of the purpose
 - the purpose of the study
 - the objectives

The protocol is drawn up before the study begins
we describe everything we do
so that anyone reading can reproduce the study exactly

Study protocol

- It is written according to certain **writing standards**
- experimental studies:
 - templates provided by the agency:
 - European Medicines Agency (EMA)
- The entire team will read and review the protocol
- The protocol is submitted to an ethics committee!!!
- After validation, it cannot be changed
 - experimental studies (all)
 - their registration in the clinical trials register:
 - EU Clinical Trials Register

How a study protocol is
made?

A. Study types

Study type. Classification 1. Research domain

Choosing a research domain

- Description of a new health phenomenon

- Evaluation of a diagnostic procedure

- Evaluation of a therapeutic procedure

- Evaluation of a risk/ protective or prognostic factors

Description of a new health phenomenon

studies that describe

- a case
- or
- a series of cases



Evaluation of a diagnostic procedure

- Some disease need a new diagnostic procedure
 - compare the new diagnostic procedure with the older one

	Disease+	Disease-	Total
New test+	a	b	
New test -	c	d	
Total			

Sensibility, specificity, positive predictive value, negative predictive value

Evaluation of a therapeutic procedure

- therapy versus placebo
- or
- new therapy versus old therapy

Predictive factor

if present = high probability of having a positive response or lack of response to a particular therapy.

Ex. radiotherapy is associated with remission of prostate cancer

Ex. erythromycin is associated with improvement of oxygen saturation in case of pneumonia diagnosis

Evaluation of a risk/ protective or prognostic

Risk factor

if present = high probability of disease

Ex. pollution influences the presence of asthma in children,

Ex. the presence of nitrogen in drinking water influences fertility, etc.

Protective factor

if present = low probability of disease

Ex. physical activity prevents obesity,

Ex. sun exposure reduces the frequency of fractures, etc.

Prognostic factor

if present = high probability of recovery or disease

Ex. old age influences tumor recurrence,

Ex. physical activity influences maintenance of normal weight after gastric sleeve surgery/stomach reduction

Study type

- Classification 2. descriptiv or analytical
- Classification 3. observational or experimental
- Classification 4. cross-sectional or longitudinal
- Classification 5. exhaustive or sampling
- Classification 6. primary or secondary

Study type. Classification 2. descriptiv or analytical

Descriptive

Discribing a case or a series of cases

Ex. a boy who ingest soy sauce in very high quatity

Ex. Covid-19 first series of 1015 cases

Assessment of disease-related indices

Ex. incidence (new case in population)

Ex. prevalence (all cases in a population) exposure statistics to factors

Ex. prevalence of periodontitis

Analytical

- Comparisons are made
- Connections are traced

Characteristics:

uses statistical tests

inferential statistical methods

e.g.

Ex. evaluating a link between butter consumption and heart attack

Ex. comparing aspirin and placebo to reduce heart attack

Ex. comparing X-rays to CT scans in cancer diagnosis

Classification 3. according to the researcher's attitude towards the study subjects



observational versus experimental

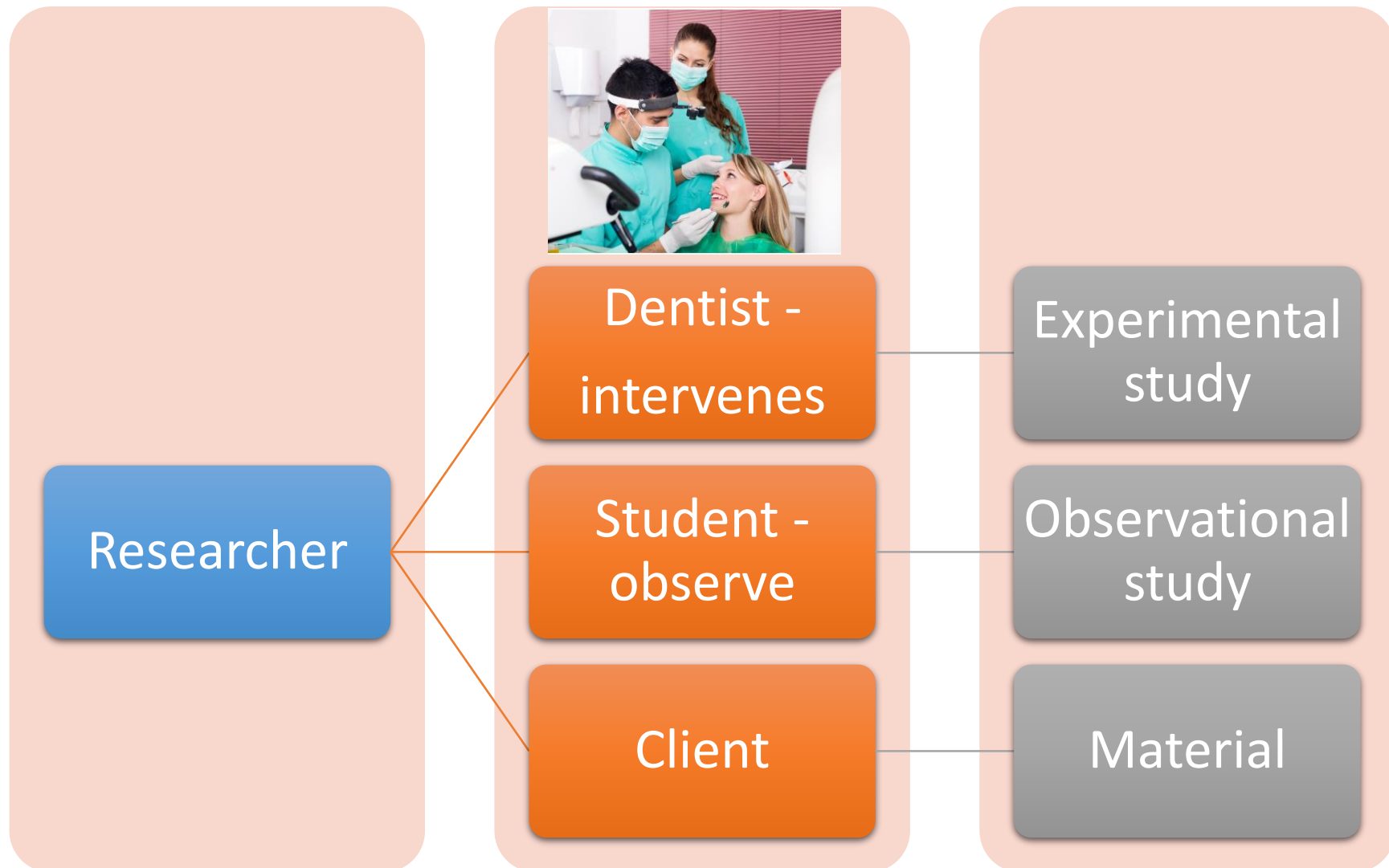
experimental if

there is direct intervention by the researcher with implications on the event of interest

Experiment
or
observation
?



Who does the intervention?



Study

Observational

researchers do not intervene
neither on the subjects,
nor on the evolution of the disease

Ex. evaluate the relationship between obesity and hypertension

Ex. compare subjects with or without heart attack to see the relationship with butter consumption

Ex. evaluating the consumption of different foods or physical effort until the onset of the infarction

Experimental

researchers intervene
on subjects
on the evolution of the disease

by

administering treatments (aspirin vs. placebo),
surgical interventions (appendectomy),
various procedures, etc.

Characteristics:

rigorously controlled
suitable for inferring causality

in humans,

mainly clinical studies
randomized controlled trial

Study

Observational

Advantages

- easy to do,
- low cost

Disadvantages

- we cannot prove causality

Experimental

Advantages

- we can demonstrate causality
- the strongest study in the hierarchy of studies

Disadvantages

- difficult to organize
- ethical implications
- high costs

Study type. Classification 4. Study duration

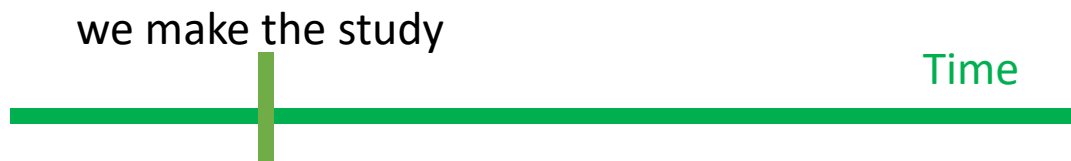
Cross-sectional

Features:

We observe subjects only once
Quick access to information

Ex. evaluate the relationship
between obesity and hypertension
at a given time

Ex. evaluate the frequency of
obesity in school children



Longitudinal

Definition:

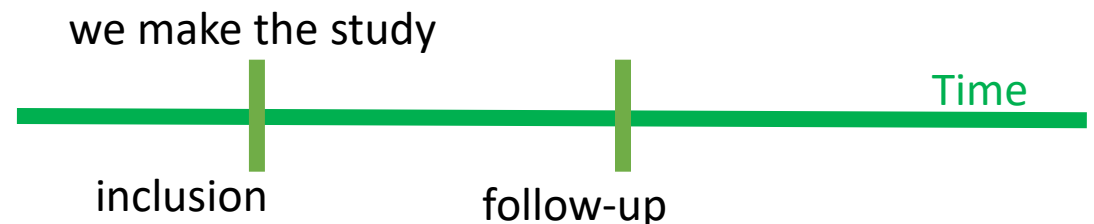
collects information about subjects at
multiple points in time

Characteristics:

organization is more difficult
access to information is longer
cost can be significant

Ex: evaluates the relationship between excessive coffee
consumption (decades) and the onset of osteoporosis

Ex. Evaluates the relationship between physical exertion
and weight



Study duration

Cross-sectional

Advantages

- prevalence can be determined
- We can study more diseases
- easier to organize
- lower cost
- shorter duration

Disadvantages

- we cannot observe whether a factor precedes an outcome
- we cannot calculate incidence or relative risk (RIE, RIN, RR, RA)
- those who die are lost from the study

Longitudinal

Advantages

- better information compared to cross-sectional studies

Disadvantages

- more difficult to organize
- more expensive

Direction over time

Longitudinal retrospective

Definition:

observations/measurements of characteristics were made in the past (before the study began)

information collected from:

observation sheets
databases

Ex. coffee consumption, smoking, food consumed

Longitudinal prospective

Definition:

observations/measurements of characteristics are made after the study begins (we have no past information)

the researcher observes/measures the characteristics of interest directly

Ex. the relationship between physical effort and anxiety reduction

Longitudinal retrospective

Advantages

- applicable to rare diseases and long incubation
- easy to organize
- low cost
- short duration

Disadvantages

- risk of observation error
- risk of recall error
- we cannot calculate incidence or relative risk (RIE, RIN, RR, RA)

Longitudinal prospective

Advantages

- more precise information
- we can calculate incidence or relative risk (RIE, RIN, RR, RA)

Disadvantages

- requires a lot of personnel
- people lost from the study (risk of attrition)
- long duration
- possible change over time of diagnostic criteria
- influence of exposure factor
- difficult to organize, expensive

Study type. Classification 5. Exhaustive versus sampling

Exhaustive (all the population)

Definition:

the entire population is studied.

Advantages

perfect representativeness
accurate results

Disadvantages

takes a long time
difficult to organize
errors
 multiple investigators,
 high ammount of data
expensive

Sampling

Definition:

a sample (several) is studied
subjects are drawn from the target
population

Advantages

easy to organize
low cost
short duration

Disadvantages

the result is an estimate
there is a risk of error
 we may not be able to generalize the results
 well

Study type. Classification 6. Primary versus secondary

Primary

Definition:

involve collecting data directly from the source

experiments

surveys

Secondary

summarising

critiquing

past studies

old published data

generate conclusions

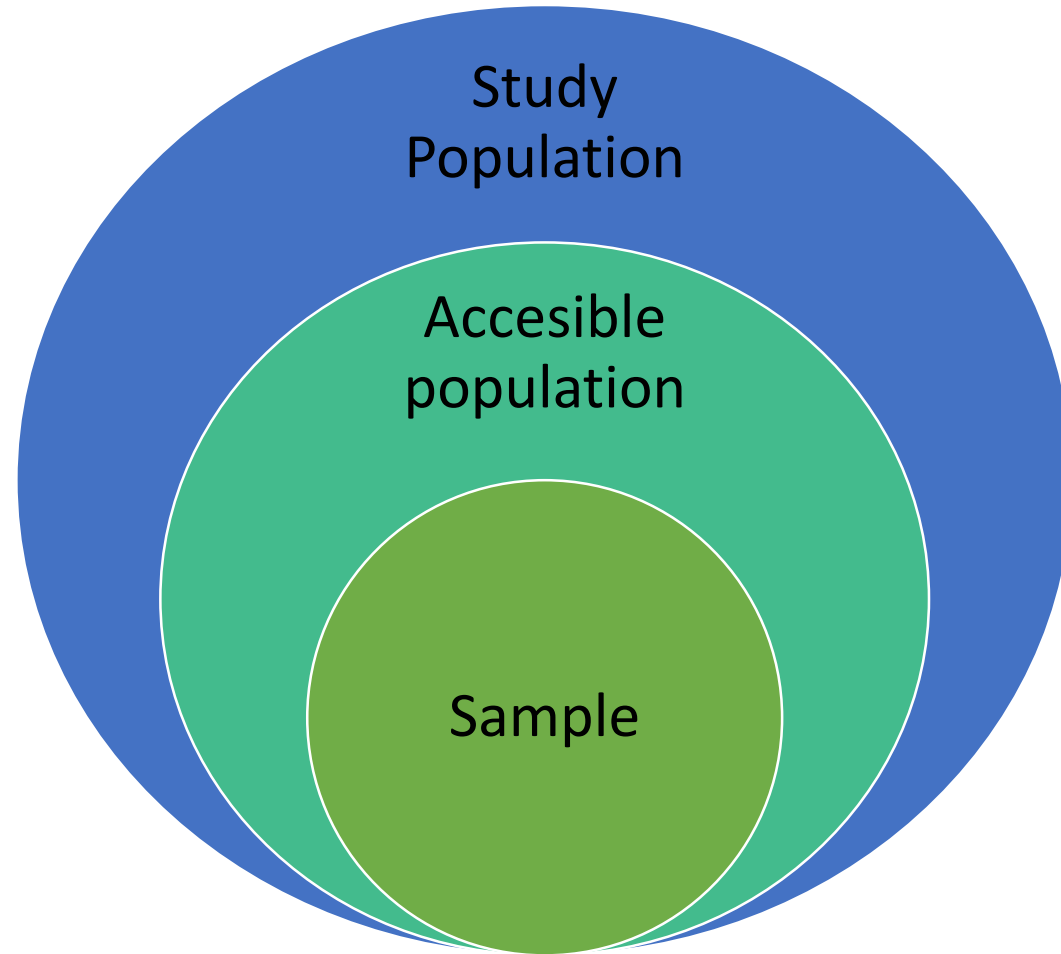
B. Sampling - population

Population

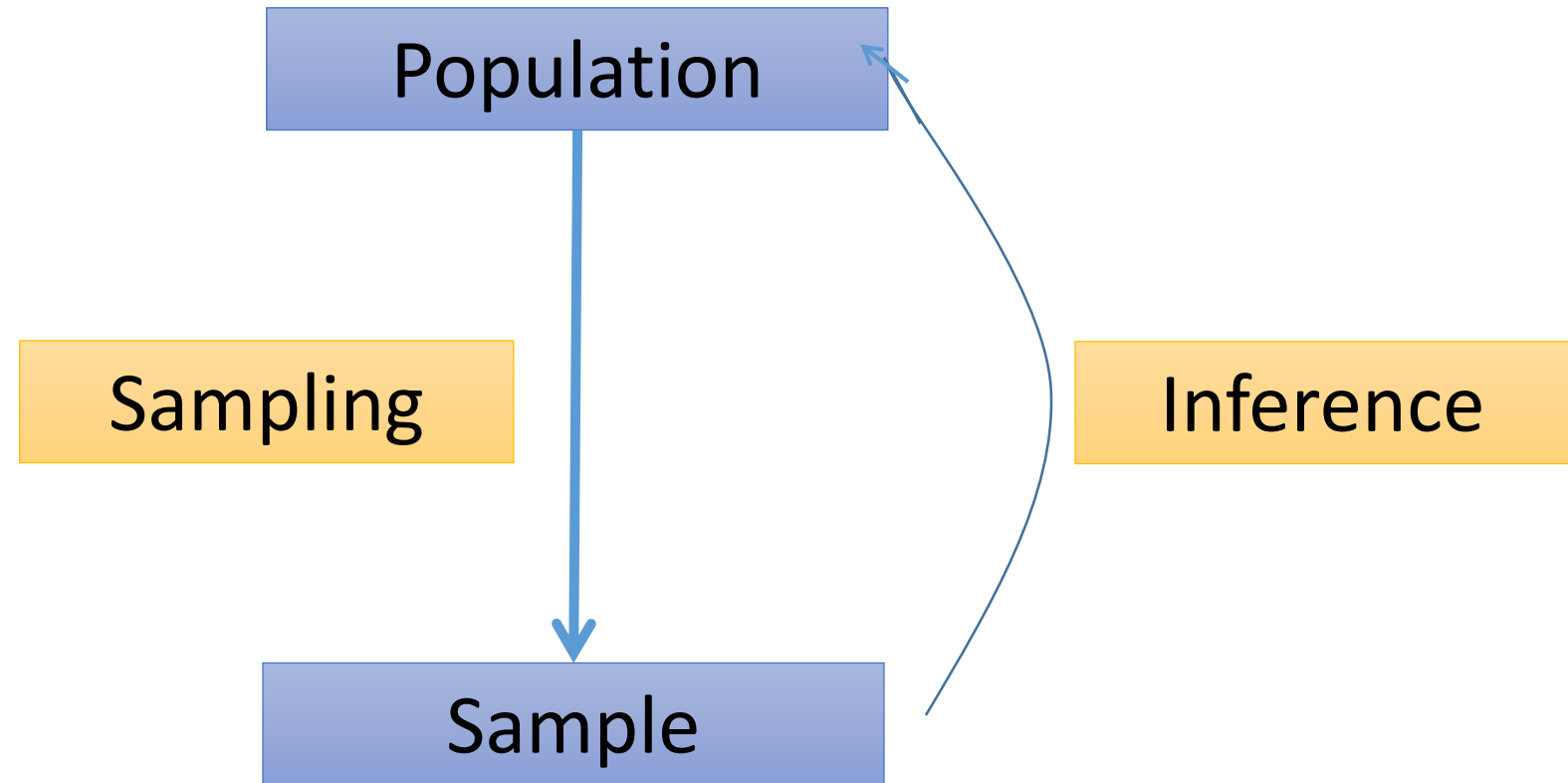
- Study population → Who is the question about?
- Accessible population
- Sample



Sampling



Research



Condition for a good estimation (inference)

Random sampling



Why random selection?

- Reduces the selection bias
- The **representative sample** for the population
 - should have the same distribution of main characteristics as the population
- where:
 - Important characteristics = in association with those studied

Sampling methods

- **Random sampling** – any of the subjects have the same chances to be selected
- **Esantionare sistematica** – fiecare al k –lea subiect este selectat
- **Esantionare stratificata** – populatia este divizata in subgrupuri si din fiecare subgrup se face selecție aleatoare

Sampling methods



probabilistic

- **random**

- any of the subjects can be selected

- **systematic**

- every kth subject is selected

- **stratified**

- the population is divided into subgroups and a random selection is made from each subgroup

- **clustered**

- the population is divided into geographical areas and a random selection is made from each subgroup

non-probabilistic – not good

Sampling methods

- random
- randomizer.org:

RESULTS

PRINTDOWNLOADCLOSE

1 Set of 2,000 Unique Numbers

Range: From 1 to 2,000— Sorted from Least to Greatest

Set #1

p1=1, p2=2, p3=3, p4=4, p5=5, p6=6, p7=7, p8=8, p9=9, p10=10, p11=11, p12=12, p13=13, p14=14, p15=15, p16=16, p17=17, p18=18, p19=19, p20=20, p21=21, p22=22, p23=23, p24=24, p25=25, p26=26, p27=27, p28=28, p29=29, p30=30, p31=31, p32=32, p33=33, p34=34, p35=35, p36=36, p37=37, p38=38, p39=39, p40=40, p41=41, p42=42, p43=43, p44=44, p45=45, p46=46, p47=47, p48=48, p49=49, p50=50, p51=51, p52=52, p53=53, p54=54, p55=55, p56=56, p57=57, p58=58, p59=59, p60=60, p61=61, p62=62, p63=63, p64=64, p65=65, p66=66, p67=67, p68=68, p69=69, p70=70, p71=71, p72=72, p73=73, p74=74, p75=75, p76=76, p77=77, p78=78, p79=79, p80=80, p81=81, p82=82, p83=83, p84=84, p85=85, p86=86, p87=87, p88=88, p89=89, p90=90, p91=91, p92=92, p93=93, p94=94, p95=95, p96=96, p97=97, p98=98, p99=99, p100=100, p101=101, p102=102, p103=103, p104=104, p105=105, p106=106, p107=107, p108=108, p109=109, p110=110, p111=111,

GENERATE NUMBERS

How many sets of numbers do you want to generate?

1

Help

How many numbers per set?

2000

Help

Number range (e.g., 1-50)

1

2000

Help

Do you wish each number in a set to remain unique?

Yes

Help

Do you wish to sort the numbers that are generated?

Yes, least to greatest

Help

How do you wish to view your random numbers?

Place Markers Within

Help

RANDOMIZE NOW!

Sample

Determining

- the number of samples

 - how many are needed?

- sampling methods

- inclusion/exclusion criteria

- sample size** (number of individuals)

Example

we want to see

- if there is a difference in exposure to obesity between those who have osteoporosis and those who do not
- from the **literature** we know that
 - those with osteoporosis are 50% obese
 - the general population is 30% obese
 - 20% difference to be demonstrated



Determining the required sample size for comparing two proportions, one of 50% and the other of 20%. <http://statpages.org/proppowr.html>

Variables of interest

- qualitative?
- quantitative?

C. Defining variables

- What characteristics will be collected?
- How will they be collected?
 - Ex.
 - we note: Obesity as obese, overweight or normal weight
 - or
 - we will measure weight and height and calculate body mass index
- Establish:
 - Questionnaires used,
 - Data collection sheets
 - How will the data be coded
 - What units of measurement will they be recorded with
 - If it is about the patient's past – specifying the date or time interval to which the information in the questionnaires refers

difference to be
demonstrated=20%

Significance Level (alpha):	0.05	(Usually 0.05) level of error =5%
Power (% chance of detecting):	80	(Usually 80) study power =80%
Group 1 Population Proportion:	.30	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

equal number of individuals in each group

Compute

Results:

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	93	93	186
With Continuity Correction:	103	103	206

if difference=10%

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Compute

Rezultate:

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	387	387	775
With Continuity Correction:	407	407	814

Dacă avem diferență de 3%?

if difference=3%

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.47	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	4355	4355	8711
With Continuity Correction:	4422	4422	8844

The smaller the difference, the greater the sample size needed to demonstrate what we set out to do.

Power of the study

Study
power = 80%

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.60	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	97	97	194
With Continuity Correction:	107	107	213

Study power = 85%

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	85	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.60	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Compute

Sample Size Required			
	Group 1	Group 2	Total
"Classical" Calculation:	111	111	221
With Continuity Correction:	120	120	241

Increase the power of study, increase the required size

Example

we want to see

- if a change in diet decrease the total cholesterol?
 - from the literature we know
 - that those with hypercholesterolemia have an average of 230 cholesterol
- we want it to decrease by at least 20 to decide that the diet is effective
 - 20 difference to demonstrate
- standard deviations: 26, respectively 33



Determining the necessary sample size to compare two means, one of 230 and the other of 210 <http://sampsiz.sourceforge.net/iface/s2.html#nm>

Assumptions:

level of error =5%	alpha =	5 (two-sided)
study power =90%	power =	90
means difference = 20	m1 =	230
	m2 =	210
standard deviation 1	sd1 =	26
standard deviation 2	sd2 =	33
equal number of individuals in each group	n2/n1 =	1

Results:

Estimated sample size:

n1 =	47
n2 =	47

means differences = 230-
220=10

Assumptions:

```
alpha =          5 (two-sided)
power  =          90
m1     =          230
m2     =          220
sd1    =          26
sd2    =          33
n2/n1  =           1
```

Estimated sample size:

```
n1 =          186
n2 =          186
```

means differences = 230-
225=5

Assumptions:

```
alpha =          5 (two-sided)
power  =          90
m1     =          230
m2     =          225
sd1    =          26
sd2    =          33
n2/n1  =           1
```

Estimated sample size:

```
n1 =          742
n2 =          742
```

The smaller the difference, the greater the effort needed to demonstrate what we set out to do.

Sample size (number of individuals)

!!! the number of individuals is determined before the study

- depends on
 - what we want to demonstrate (means, frequencies)
 - how many samples? (groups of patients with the same characteristic)
 - how precise we want the study to be
 - the size of the confidence interval
 - large (not very precise)
 - narrow (precise)
 - how big we expect the difference between the groups to be
 - big
 - small

Sample size (number of individuals)

depends on

- how big we expect the difference between groups to be

 - frequencies

 - in case of exposure to the risk factor

 - in case of disease

 - average

 - cholesterol before and after treatment

what the population looks like:

- standard deviation

 - very diverse/not very diverse

D. Mention the risks of errors

- Identifying possible errors
- Identifying ways to control them
 - Ex. protocol related to incomplete data
 - Ex. protocol related to patients lost from the study

E. Standardization of methods

- Defining measurement / observation / recording methods
 - Understandable, clear
 - Feasible (can be done)
 - Accurate
 - Reproducible – anyone can reproduce the study
- As for the equipment / laboratory
 - it is preferable to be one throughout the study

F. Establishing the data analysis plan

- Database creation
- Database validation
- Transfer method
- Persons who have access to the database
 - levels of accessibility
- Statistical methods that will be applied
- What indicators will be calculated

G. Other aspects

- Staff – training
- Financial side: will the participant be paid?
- Will the staff be paid?
- Where will the funds come from?
- Ethical considerations
 - Medical ethics rules
- Data protection
- Establishing the duration of each stage

Summary of the previous course

Researcher's attitude

- observational
- experimental

Objective

- descriptive
- analytic

Selection method

- sample
- exhaustive

Selection method

- sample
- exhaustive

Data collection duration

- cross-sectional study
- longitudinal study
 - prospective
 - retrospective

Research areas

- **Description of a health phenomenon**
 - descriptive studies
- **Evaluation of a diagnostic procedure**
 - some conditions require a new diagnostic method
- **Evaluation of a therapeutic approach**
 - study of the efficiency/safety of new medical treatments or procedures
- **Research on prognostic factors (risk, protective)**
 - study of the association between a risk factor and disease

Thank you!