FROM SAMPLE TO DATA COLLECTION

2 OBJECTIVES

- Choosing the sample type
- Computing the necessary sample size
- Clasification of the research by the type of the intervention
- Clasification of the research by the data collection method
- Variables

3

- Do I need a sample?
- Which method should I use to choose the subjects?
- How many subjects must be in a sample?
- What procedure do I have to use?
- Which are the (main) characteristics of the subjects to be collected?

4 EXAMPLE

- **You are studying the usefulness of wood crutches vs. aluminum crutches for people with unilateral amputation.**

5 EXAMPLE

- **Research question: Which device (wood crutches or aluminum crutches) is more useful to help people with unilateral amputation to walk?**

6 EXAMPLE

- Putting in practice an experiment which will help you to decide **which device (wood crutches or aluminum crutches) is more useful to help people with unilateral amputation to walk.**

7 DO I NEED A SAMPLE

SAMPLE SELECTION

- **Objectives**
- **Type of the study**
- **Available time**
- **Financial and human resources which are available**
- **Procedure**
- **Data access**

9 MAIN TYPES OF MEDICAL RESEARCH

- Descriptive
- Risk / Protective factors
- Diagnostic signs
- The clinical trial
- Syntheses
- Systematic syntheses
- Meta-analysis
- Experimental research

SAMPLE SELECTION – OUR STUDY

OUR EXAMPLE

- **You are studying the usefulness of wood crutches or aluminum crutches for people with unilateral amputation.**
- **Your coordinator works on the Recovery Hospital**
- **Target population: People with unilateral amputation**
- **Accessible population: People with unilateral amputation using the services offered by the Recovery Hospital**

SAMPLE TYPES

Random sample:

- **Random extraction of elements from the study population**
- **Each element has the same chance of being included in the sample**
- **RANDBETWEEN**

13 SOMETIMES...

- Convenience sample
- Snowball sample
- ...

EXAMPLE – CONVENIENCE SAMPLE

- Title: The usefulness of virtual patients in the study of thyroid pathology

- Target population: medical students
- Available population: Students on UMF Iuliu Haţieganu
- Sample: Students (friends of the author) who use the virtual patient

15 EXAMPLE – SNOWBALL SAMPLE

- Challenges being a member of the LGBT community and also medical student
- Target population: Medical students, members of LGBT community
- Available population: ?
- Sample: - snowball – start from few (known by the author) members of local (?) LGBT community and asking them to bring in the sample other members of LGBT community (not necessary from local community)

16 OUR EXAMPLE

- **The usefulness of wood crutches or aluminum crutches for people with unilateral amputation.**

- Consecutive patients with unilateral amputation using wood crutch or aluminum crutch who accept to join the study.

COMPUTING THE SAMPLE SIZE

EXAMPLE – FOR DESCRIPTIVE STUDIES

18

- It is desired to evaluate the prevalence of a disease in a certain population. It is known that in other studies this prevalence was of the order of 30%(=f). If we want a precision of the prevalence of k=4% , for a risk level α = 0.05 (Z_α = 1.96) the estimation must be made using a sample containing more than: n = 505 persons

- .

$$n = \frac{Z_\alpha^2 \cdot f(1-f)}{k^2}$$

$$\frac{(1.96)^2 \times 0.3 \times 0.7}{(0.04)^2}$$

19 COMPUTING THE SAMPLE SIZE

- When we want to compare the average from the sample with the average in population

$$n = \left[z\frac{\alpha}{2} \frac{\sigma}{E} \right]^2 \quad ; \; for\ \alpha = 0{,}05 \quad z\frac{\alpha}{2} = 1{,}96 \quad n = \left[1{,}96 \frac{\sigma}{E} \right]^2$$

HOW CAN WE KNOW THE STANDARD DEVIATION?

- Other studies
- Pilot study

20 EXAMPLE

- The dean of the faculty wants to check to what extent students' assessments for a teacher are correct (which is the minimum sample required for the estimator to be statistically correct)
- The required 95% confidence level, the error should not exceed 0.25 points
- From the evaluation of the answers, the standard deviation was 1.5
- $N_{0,25} = (1,96)^2 * (1,5)^2 / (0,25)^2 = 138$

21 EXAMPLE

- We want to evaluate the effect of high voltage lines in the occurrence of cardiovascular diseases. From descriptive studies, the odds to develop cardiovascular diseases if living next to high voltage lines is twice as usual.

22 RESULTS…

- https://epitools.ausvet.com.au/casecontrolss

EPITOOLS

Sample size for a Case-control study

Analysed: Wed Jan 10, 2024 @ 15:08 UTC

Results

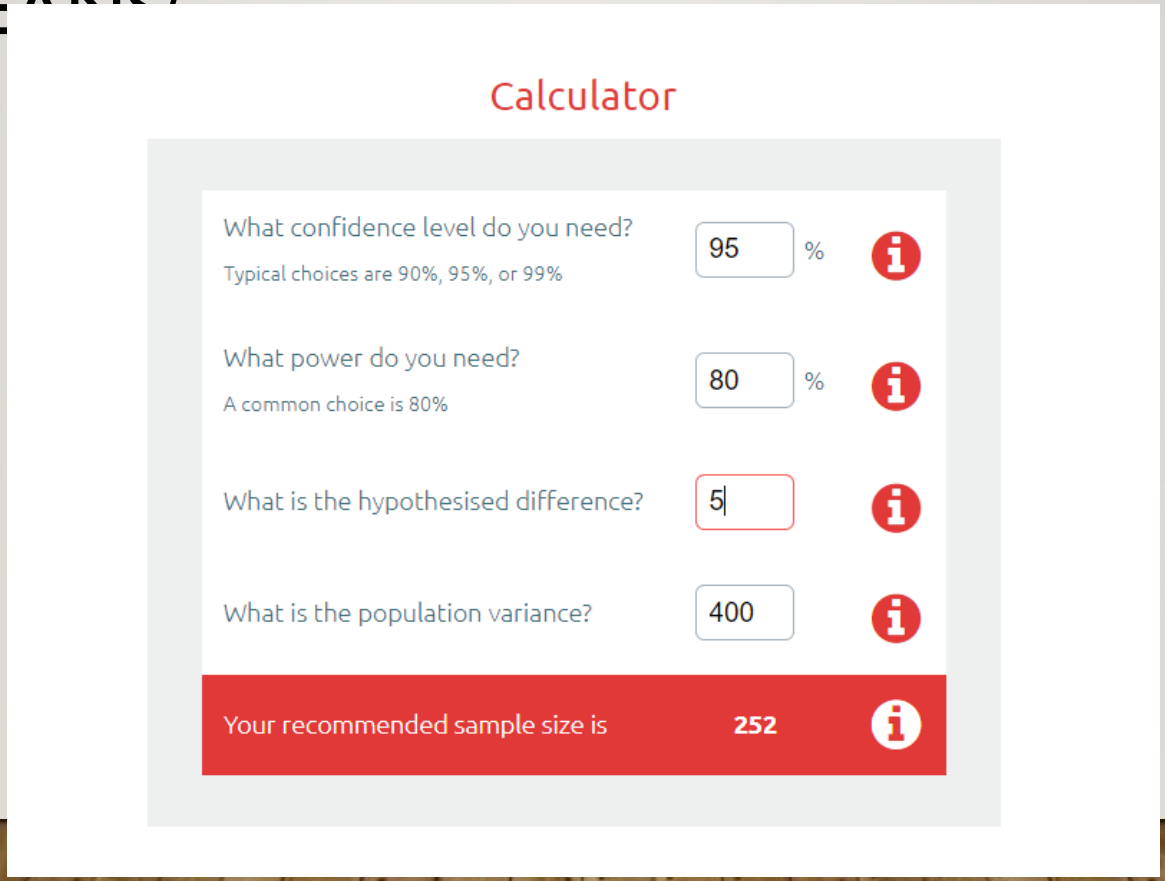
Sample size for specified values

Expected proportion in controls	0.1
Assumed odds ratio	2
Confidence level	0.95
Power	0.8
Study type	Case-control study
Sample size per group	280
Total sample size (both groups):	560

23 OUR EXAMPLE

- The usefulness of **wood crutches or aluminum crutches for people with unilateral amputation.**
- **The variable: time in seconds to walk 100 m**
- **We will consider a type of crutch more effective if the time to walk 100 m helped by that type of crutch is, on average, with 5' less that time to walk 100 m helped by the other type of crutch**
- **The common variation of the time needed to walk 100 m for people with unilateral amputation is 400'**

24 HTTPS://SELECT-STATISTICS.CO.UK/CALCULATORS/SAMPLE-SIZE-CALCULATOR-TWO-MEANS/



25 WHAT IF WE DON'T HAVE WHERE TO COLLECT SO MANY CASES?

- **Magic number…**

30

26 EXAMPLE

- The usefulness of **wood crutches or aluminum crutches for people with unilateral amputation.**

- **Sample: 30 consecutive patients with unilateral amputation which come to the Recovery Hospital for treatment and accept to be part of the study, having wood crutches and 30 consecutive patients with unilateral amputation which come to the Recovery Hospital for treatment and accept to be part of the study, having aluminum crutches**

27 PROCEDURE

CLASSIFICATION OF STUDIES BY DATA COLLECTION

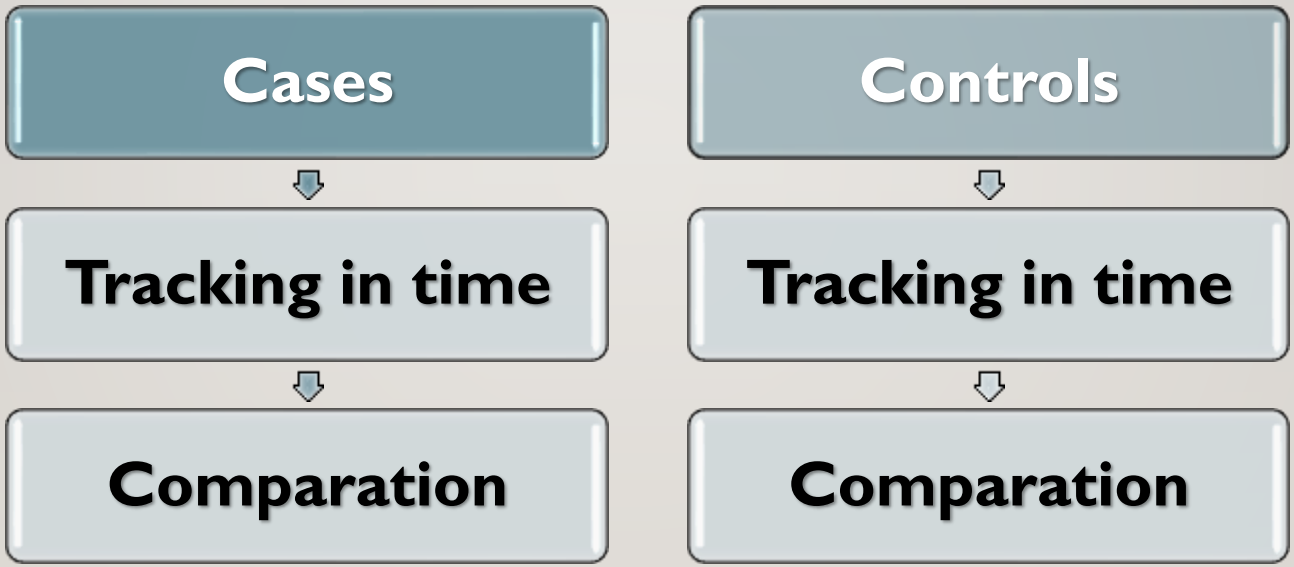
29 DURATION OF DATA COLLECTION

- Transversal studies

LONGITUDINAL RESEARCH

Prospective

Retrospective



TYPES OF DATA COLLECTION

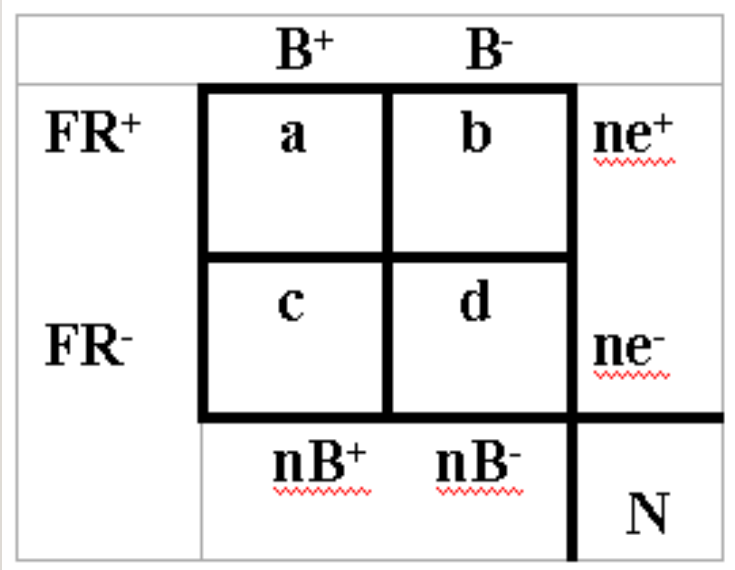
1. Representative sample
2. Exposed non-exposed
3. Case control

REPRESENTATIVE SAMPLE

- **The group of subjects studied is representative of the target population**
- **To be used:**
  - **descriptive study (description of a health phenomenon)**
  - **analytical study (in which the comparison of at least two groups of subjects is mandatory).**

REPRESENTATIVE SAMPLE

- Total sample size= N



REPRESENTATIVE SAMPLE

- **Allows most types of calculations can be performed, with no sampling restrictions:**
  - **calculating the prevalence of the disease**
  - **calculation of exposure frequency**
  - **measuring the degree of association between the incriminating factor (risk or prognosis) and the disease.**

REPRESENTATIVE SAMPLE

- **Limits:**
  - **difficulties in tracking a large number of subjects**
    - **cost issues**
    - **the risk of decreasing the subjects' interest in the study**
    - **the risk of a large number of loss of sight.**

NON-PROBABILISTIC VS REPRESENTATIVE

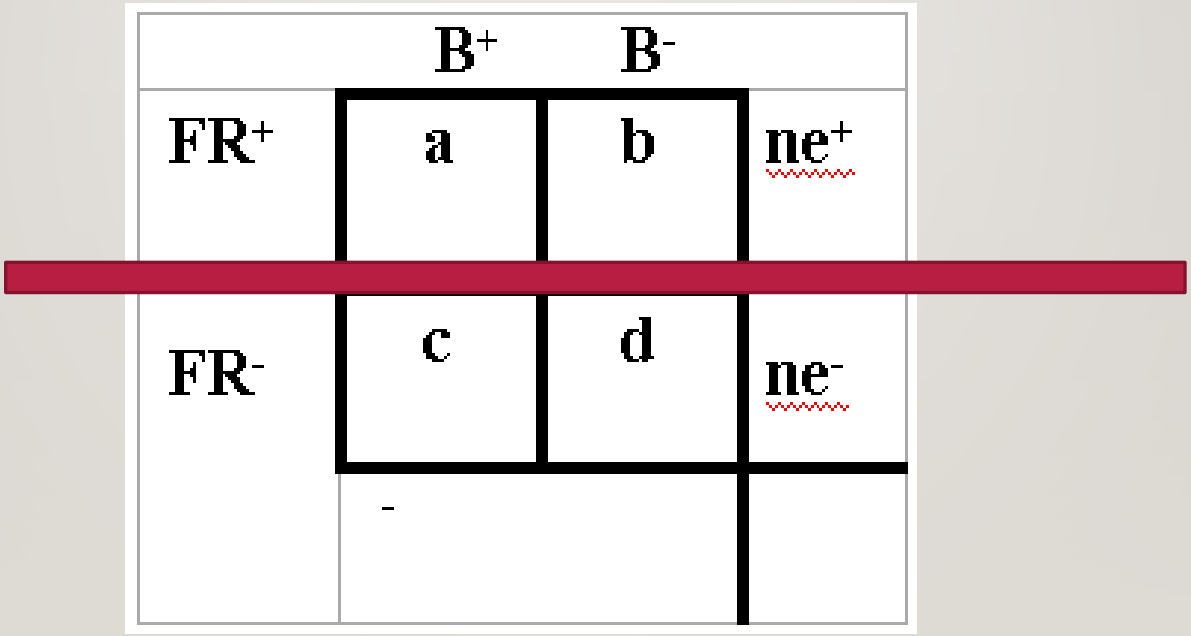
- The non-probabilistic samples the results allow the same type of calculation like the representative samples

EXPOSED-NONEXPOSED

- **Divide the subjects into two categories :**
  - **exposed**
  - **nonexposed**
- **The exposure factor is fixed, and the occurrence of the disease is studied over time**
- **Disease = random factor**

EXPOSED-NONEXPOSED

- **The number of subjects in the two groups may be the same or different.**

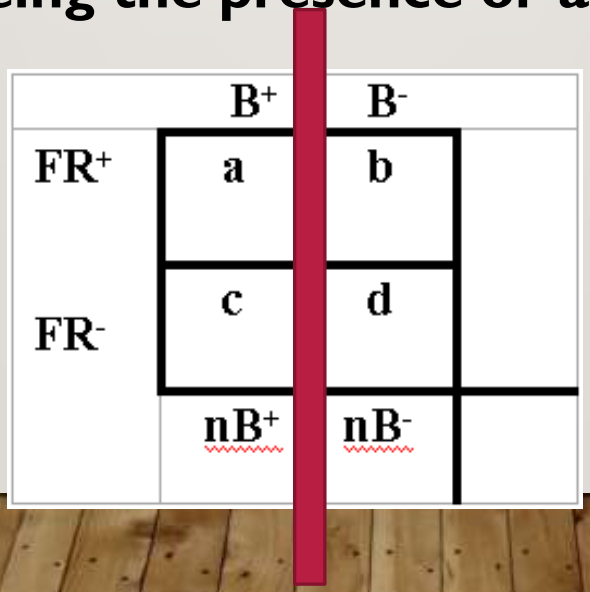


YOU CAN COMPUTE:

- RR – Risk ratio
- RD – Risk Difference
- …

CASE - CONTROL

- predefining two groups of subjects:
  - B + = sick = case group
  - B- = disease free = control group
  - the main criterion for inclusion or exclusion in one group or another being the presence or absence of the disease



CASE - CONTROL

- **disease = controlled factor**
- **exposure = random factor**
  - **the disease is predetermined and the presence of exposure (risk factor / prognosis) in the onset (or evolution) of the disease is generally retrospectively studied**

OR = ODDS RATIO

Disease / Exposure	D+ Sick	D- Healthy	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	n

- $Odds_{exp}$ = [a/(a+b)]/[b/(a+b)]=a/b
- $Odds_{ne\text{-}exp}$ = [c/(c+d)]/[d/(c+d)]=c/d
- OR = (a/b)/(c/d) = ad/bc

43 YOUR THESIS MAY BE:

- Descriptive
- Risk factors
- Diagnostic signs
- The clinical trial
- Survival analysis
- Syntheses
- Systematic syntheses
- Meta-analysis

44 DESCRIPTIVE RESEARCH

- Descriptive studies should not be seen only as a complement or supplement to analytical research.
- They can also be used to generate hypotheses regarding the solutions of some problems, hypotheses that can later be tested by quantitative methods based on descriptive study results or a combination of descriptive and analytical methods.

45 RISK FACTORS

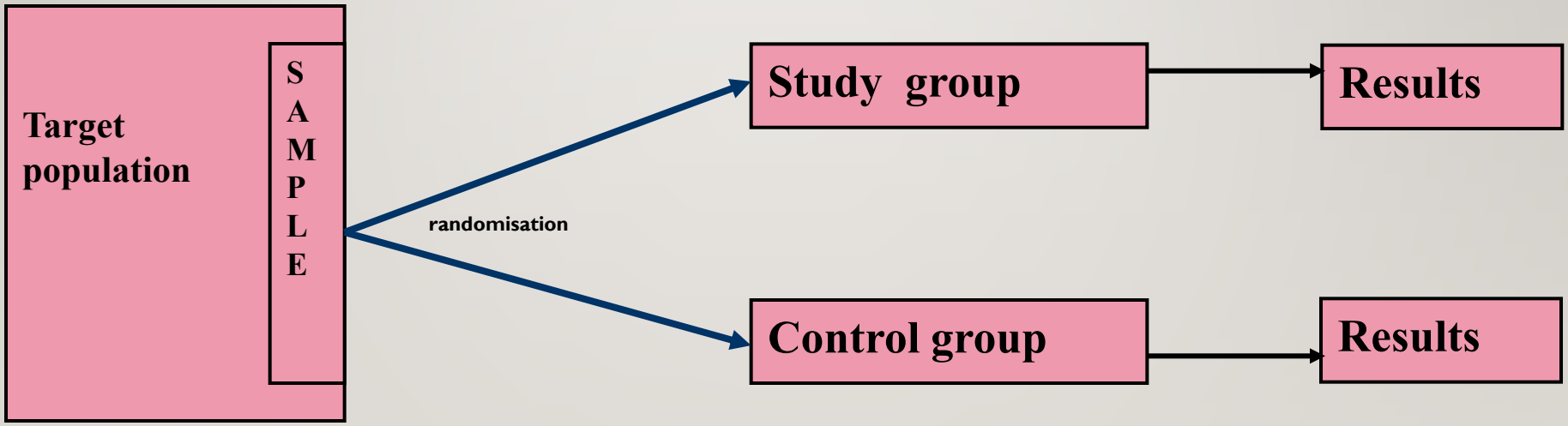
- Case - control
- Exposed – Nonexposed
- Survival

46 DIAGNOSTIC SIGNS / TESTS

- Se, Sp
- PPV, NPV

47 RANDOMIZED AND CONTROLLED CLINICAL TRIAL

- Evaluation of therapeutic interventions



48 SYNTHESES

Characteristics	Traditional synthesis	**Systematic synthesis**
The strategy used to search for primary data	**Usually limited to a single database - MEDLINE**	**Well-developed search strategy, multiple databases**
Explicit description of the search strategy	**Not included**	**Included**
Extraction of data from the primary source	**Subjective and random choice**	**Systematic evaluation of the quality of all identified items, using explicit criteria**
Analysis of results from primary sources	**The use of various techniques**	**Systematic analysis using valid methods**

49 SYSTEMATIC SYNTHESIS

- **Answer to a clinical question**
  - Comprehensive search to identify studies relevant to the question we are looking for an answer to:
    - electronic databases
    - www
    - searching for syntheses on the given topic
    - contacting experts in the field
    - searching for unpublished studies
    - "Gray literature": reports, data from the pharmaceutical industry
    - Library
    - Cochrane Library
    - Strict inclusion criteria

50 SYSTEMATIC SYNTHESIS

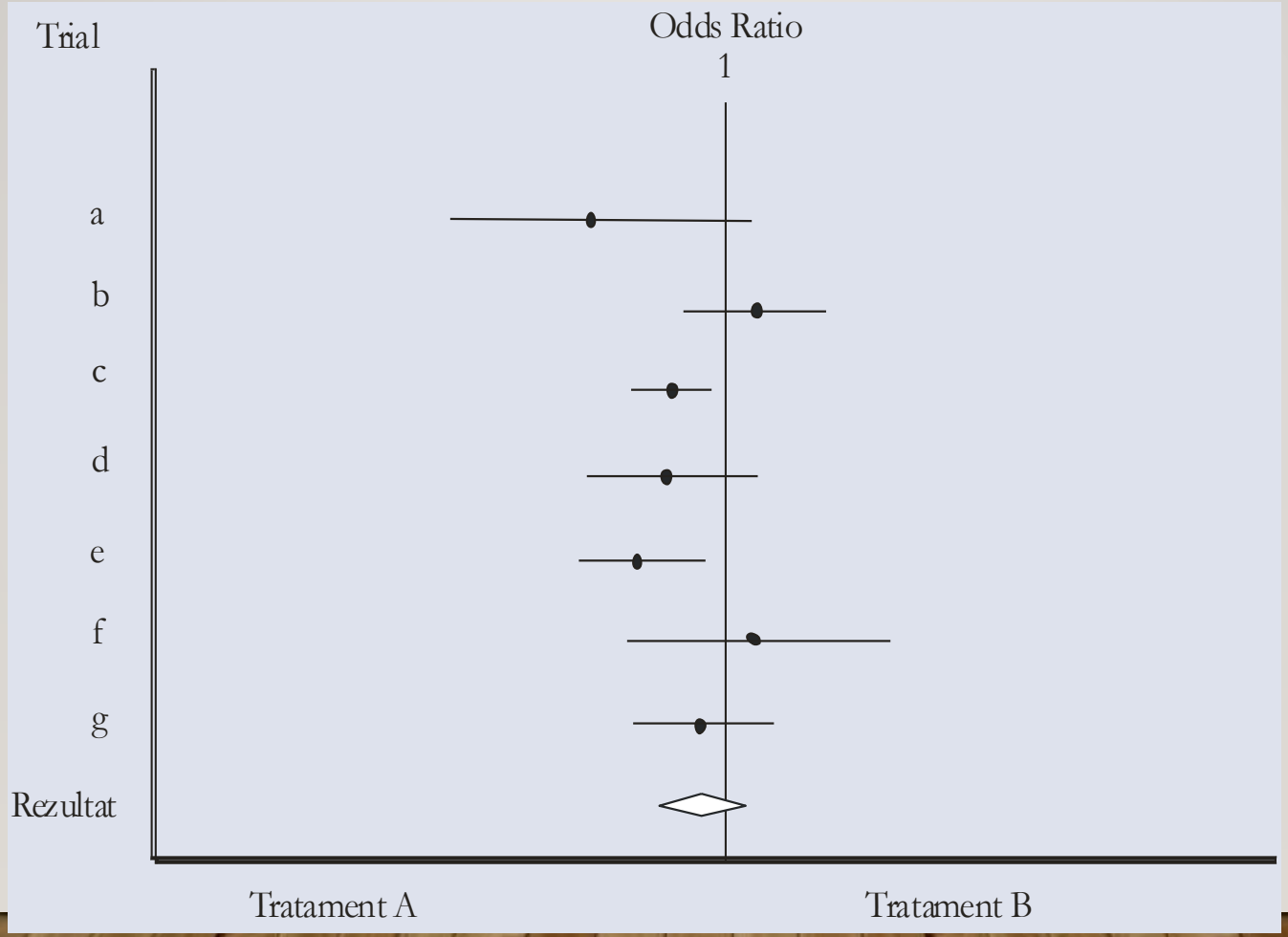
- **The results of the synthesis depend on the results of the studies included in the synthesis**

- **Can the results be applied to the patient?**

META-ANALYSIS

- The meta-analysis is a mathematical sum of several results of the primary studies, all of which used similar methods and the objectives were aimed at finding the answer to the same clinical question.
- The results of the studies included in the meta-analyzes are presented in a standard form: forest-type representation

META-ANALYSIS



Theoretical forest-type graph to illustrate the results of a meta-analysis comparing treatments A and B. The individual trials are represented by horizontal lines (confidence intervals) and black dots (centers of confidence intervals). The rhombus is the representation of all the results taken into account.

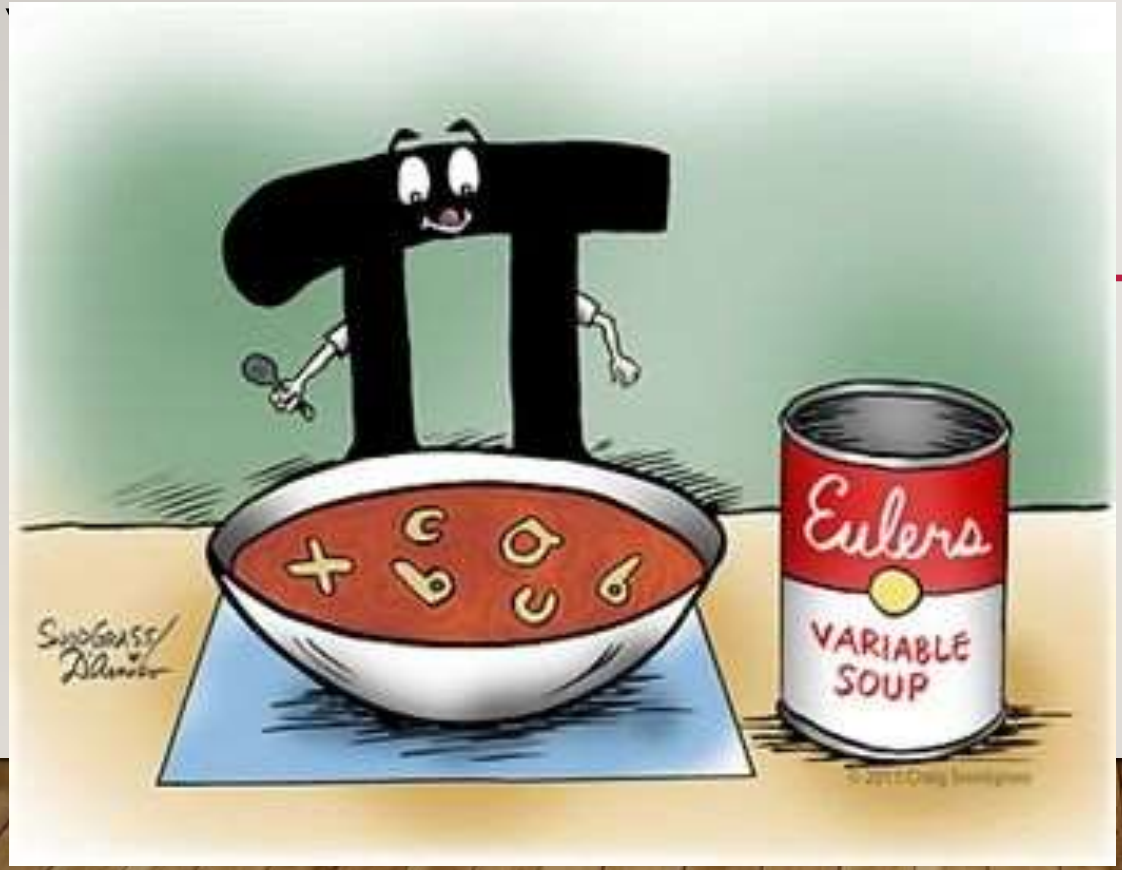
53 OUR STUDY

- **The usefulness of wood crutches or aluminum crutches for people with unilateral amputation.**

- Procedure – for all the participants we'll measure the time needed to walk 100 m, and then we'll compare the results by groups

- Transversal, convenience sample

54



55

Topic - >

Title ->

Hypothesis ->

Study Type -> Variables

IDENTIFICATION OF VARIABLES

- **The characteristics studied on the subjects in a research, which fluctuate from one individual to another individual, are called variables.**
- **The variables must be defined from the beginning of the study.**
- **Coded and recorded in tables, sheets or on magnetic media, the variables become data.**
- **Types of variables:**
  - **qualitative**
  - **quantitative**
  - **Survival**

1. THE QUALITATIVE VARIABLES

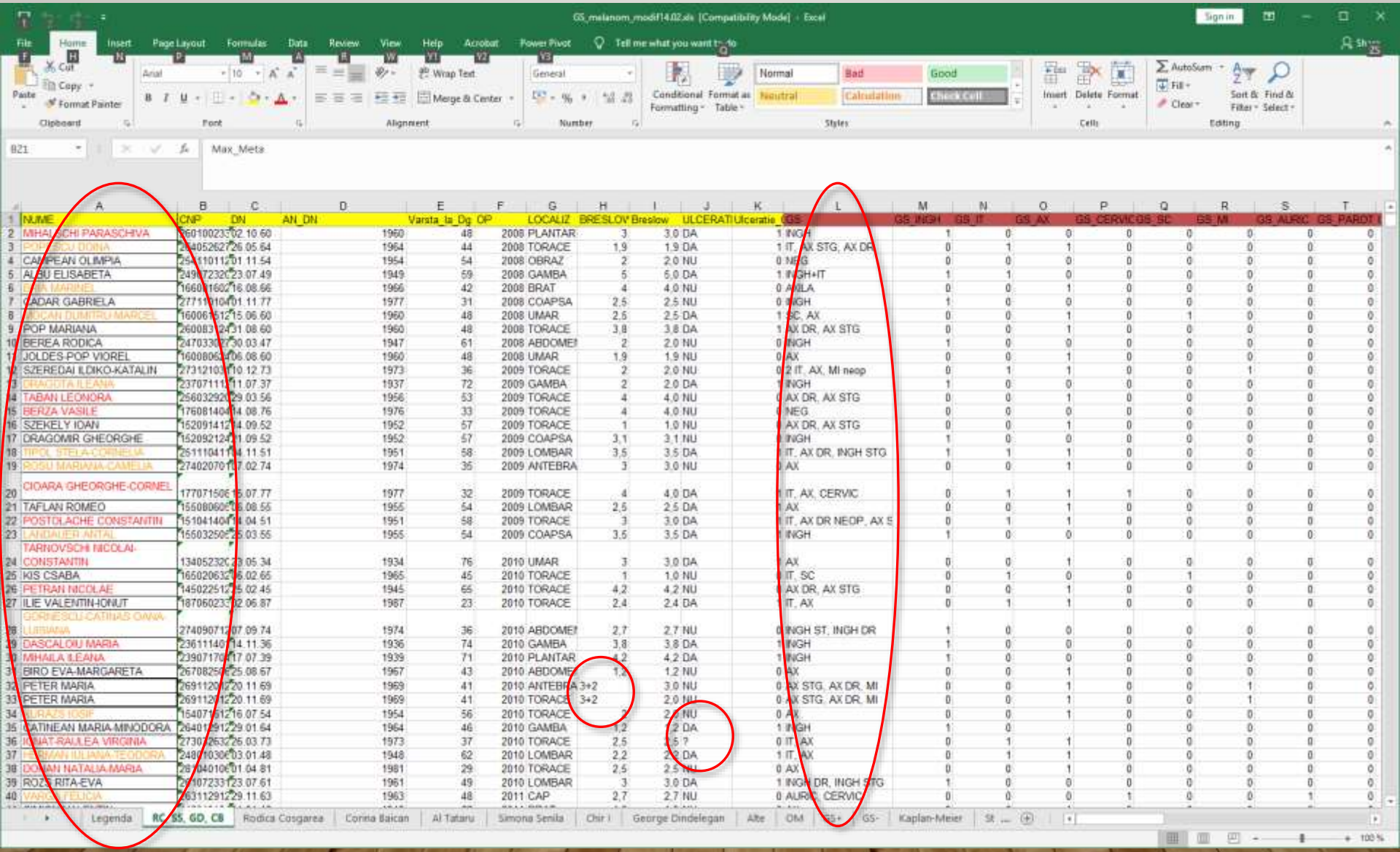
- **they are immeasurable, finite, and the calculation of the mean makes no sense**
  - **nominal - groups subjects into categories that cannot be ordered (example: eye color)**
  - **nominal ordered - subjects are grouped into categories that can be ordered (example: health = poor, better, very good)**
  - **dichotomous - subjects are always grouped into two categories (example: sick and healthy)**
- **Dichotomous variables are a special case of nominal variables.**

2. QUANTITATIVE VARIABLES

- **Can be measured**
  - continuous - the size is measurable; can take an infinity of values; the calculation of the average always has significance (example: height, weight, SBP, etc.)
  - discrete - discontinuous, - contain only integers. The value of the average does not always mean. The analysis of the results obtained based on calculations using discrete variables is delicate (example: the number of children in a family).

3. SURVIVAL VARIABLES

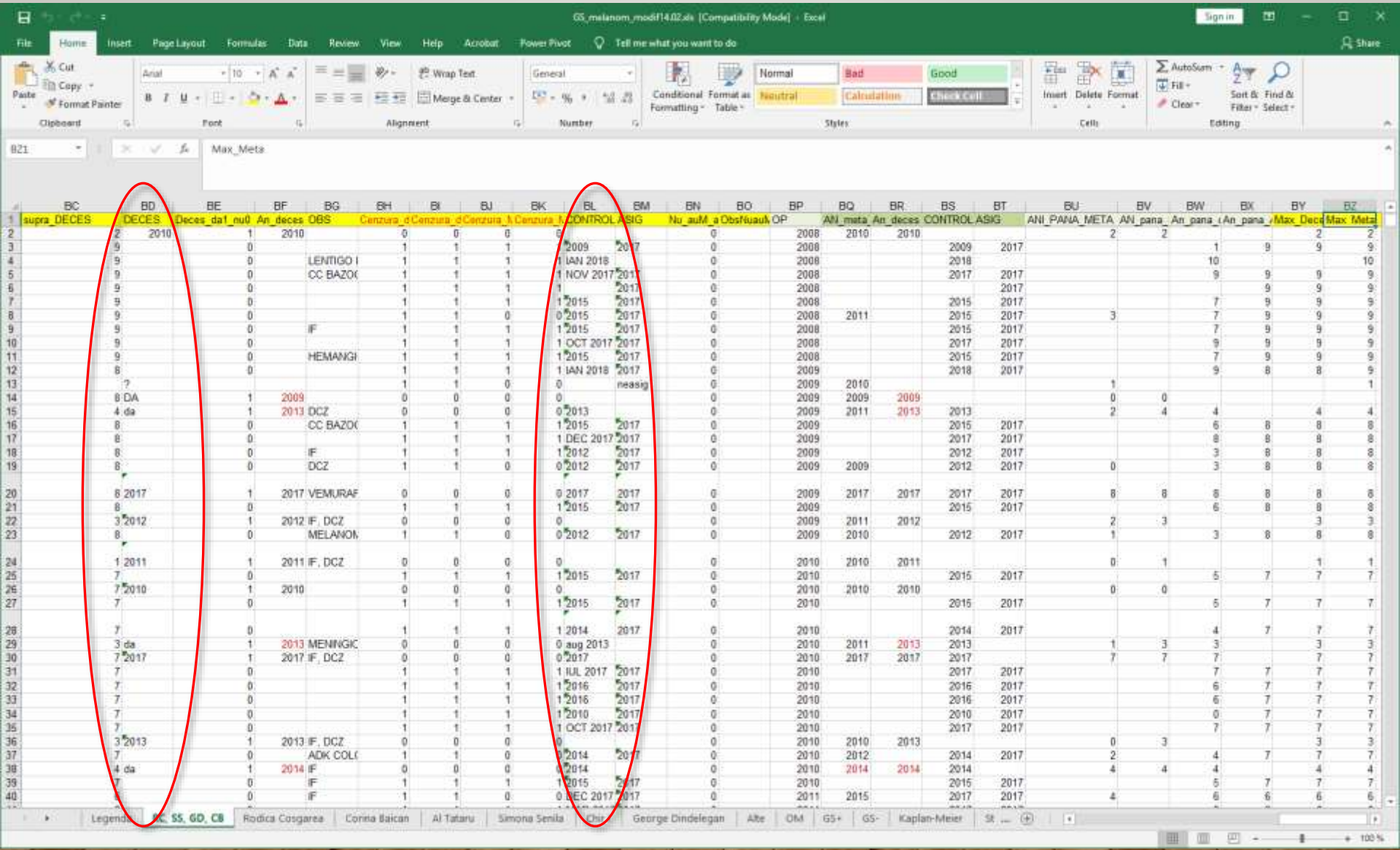
- **correspond to the time elapsed between the inclusion of a subject in a study and the appearance of a predefined element of the study:**
  - **Death**
  - **Metastasis**
  - **Complication**
  - **Symptom**
  - **Signal**



GS_melanom_modif14.02.xls [Compatibility Mode] - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	NUME	CNP	DN	AN_DN	Varsta_la_Dg	OP	LOCALIZ	BRESLOV	Breslow	ULCERATI	Ulceratie_C	GS	GS_INGH	GS_IT	GS_AX	GS_CERVIC	GS_SC	GS_MI	GS_AURIC	GS_PARDT
2	MIHALACHI PARASCHIVA	2601002330210.60		1960	48	2008	PLANTAR	3	3,0	DA	1	ING H	1	0	0	0	0	0	0	0
3	POPESCU DOINA	2640526270605.64		1964	44	2008	TORACE	1,9	1,9	DA	1	IT, AX STG, AX DR	0	1	1	0	0	0	0	0
4	CAMPEAN OLIMPIA	2541101110111.54		1954	54	2008	OBRAZ	2	2,0	NU	0	NEG	0	0	0	0	0	0	0	0
5	ALBU ELISABETA	2490723260723.07.49		1949	59	2008	GAMBA	5	5,0	DA	1	INGH+IT	1	1	0	0	0	0	0	0
6	[illegible] MARINEL	1660816021608.66		1966	42	2008	BRAT	4	4,0	NU	0	AXILA	0	0	1	0	0	0	0	0
7	CADAR GABRIELA	2771110140111.77		1977	31	2008	COAPSA	2,5	2,5	NU	0	INGH	1	0	0	0	0	0	0	0
8	[illegible]DICAN DUMITRU-MARCEL	1600615121506.60		1960	48	2008	UMAR	2,5	2,5	DA	1	SC, AX	0	0	1	0	1	0	0	0
9	POP MARIANA	2600831240108.60		1960	48	2008	TORACE	3,8	3,8	DA	1	AX DR, AX STG	0	0	1	0	0	0	0	0
10	BEREA RODICA	2470330073003.47		1947	61	2008	ABDOMEI	2	2,0	NU	0	INGH	1	0	0	0	0	0	0	0
11	JOLDES-POP VIOREL	1600806240608.60		1960	48	2008	UMAR	1,9	1,9	NU	0	AX	0	0	1	0	0	0	0	0
12	SZEREDAI ILDIKO-KATALIN	2731210331012.73		1973	36	2009	TORACE	2	2,0	NU	0	2 IT, AX, MI neop	0	1	1	0	0	1	0	0
13	DRAGOTA ILEANA	2370711110107.37		1937	72	2009	GAMBA	2	2,0	DA	1	INGH	1	0	0	0	0	0	0	0
14	TABAN LEONORA	2560329220903.56		1956	53	2009	TORACE	4	4,0	NU	0	AX DR, AX STG	0	0	1	0	0	0	0	0
15	BERZA VASILE	1760814041408.76		1976	33	2009	TORACE	4	4,0	NU	0	NEG	0	0	0	0	0	0	0	0
16	SZEKELY IOAN	1520914121409.52		1952	57	2009	TORACE	1	1,0	NU	0	AX DR, AX STG	0	0	1	0	0	0	0	0
17	DRAGOMIR GHEORGHE	1520921240109.52		1952	57	2009	COAPSA	3,1	3,1	NU	0	INGH	1	0	0	0	0	0	0	0
18	TIPOL STELA-CORNELIA	2511104110411.51		1951	58	2009	LOMBAR	3,5	3,5	DA	1	IT, AX DR, INGH STG	1	1	1	0	0	0	0	0
19	POSU MARIANA-CAMELIA	2740207010702.74		1974	35	2009	ANTEBRA	3	3,0	NU	0	AX	0	0	1	0	0	0	0	0
20	CIOARA GHEORGHE-CORNEL	1770715061607.77		1977	32	2009	TORACE	4	4,0	DA	1	IT, AX, CERVIC	0	1	1	1	0	0	0	0
21	TAFLAN ROMEO	1550806062606.55		1955	54	2009	LOMBAR	2,5	2,5	DA	1	AX	0	0	1	0	0	0	0	0
22	POSTOLACHE CONSTANTIN	1510414041404.51		1951	58	2009	TORACE	3	3,0	DA	1	IT, AX DR NEOP, AX S	0	1	1	0	0	0	0	0
23	LANDAUER ANTAL	1550325052503.55		1955	54	2009	COAPSA	3,5	3,5	DA	1	INGH	1	0	0	0	0	0	0	0
24	TARNOVSCHI NICOLAI-CONSTANTIN	1340523202305.34		1934	76	2010	UMAR	3	3,0	DA	1	AX	0	0	1	0	0	0	0	0
25	KIS CSABA	1650206320602.65		1965	45	2010	TORACE	1	1,0	NU	0	IT, SC	0	1	0	0	1	0	0	0
26	PETRAN NICOLAE	1450225122502.45		1945	65	2010	TORACE	4,2	4,2	NU	0	AX DR, AX STG	0	0	1	0	0	0	0	0
27	ILIE VALENTIN-IONUT	1870602330206.87		1987	23	2010	TORACE	2,4	2,4	DA	1	IT, AX	0	1	1	0	0	0	0	0
28	[illegible] LILIANA	2740907120709.74		1974	36	2010	ABDOMEI	2,7	2,7	NU	0	INGH ST, INGH DR	1	0	0	0	0	0	0	0
29	DASCALOIU MARIA	2361114041411.36		1936	74	2010	GAMBA	3,8	3,8	DA	1	INGH	1	0	0	0	0	0	0	0
30	MIHAILA ILEANA	2390717081707.39		1939	71	2010	PLANTAR	4,2	4,2	DA	1	INGH	1	0	0	0	0	0	0	0
31	BIRO EVA-MARGARETA	2670825062505.08.67		1967	43	2010	ABDOMEI	1,2	1,2	NU	0	AX	0	0	1	0	0	0	0	0
32	PETER MARIA	2691120072011.69		1969	41	2010	ANTEBRA	3+2	3,0	NU	0	AX STG, AX DR, MI	0	0	1	0	0	1	0	0
33	PETER MARIA	2691120072011.69		1969	41	2010	TORACE	3+2	2,0	NU	0	AX STG, AX DR, MI	0	0	1	0	0	1	0	0
34	[illegible] IOSIF	1540715121607.54		1954	56	2010	TORACE	2	2,0	NU	0	AX	0	0	1	0	0	0	0	0
35	CATINEAN MARIA-MINODORA	2640129122901.64		1964	46	2010	GAMBA	1,2	1,2	DA	1	INGH	1	0	0	0	0	0	0	0
36	IGNAT-RAILEA VIRGINIA	2730726262603.73		1973	37	2010	TORACE	2,5	2,5	?	0	IT, AX	0	1	1	0	0	0	0	0
37	[illegible]RMAN IULIANA-TEODORA	2480810300301.48		1948	62	2010	LOMBAR	2,2	2,2	DA	1	IT, AX	0	1	1	0	0	0	0	0
38	DOMAN NATALIA-MARIA	2810401060104.81		1981	29	2010	TORACE	2,5	2,5	NU	0	AX	0	0	1	0	0	0	0	0
39	ROZS RITA-EVA	2610723312307.61		1961	49	2010	LOMBAR	3	3,0	DA	1	INGH DR, INGH STG	1	0	0	0	0	0	0	0
40	VARGA FELICIA	2631129122911.63		1963	48	2011	CAP	2,7	2,7	NU	0	AURIC, CERVIC	0	0	0	1	0	0	1	0

Legenda | RC, SS, GD, CB | Rodica Cosgarea | Corina Baican | Al Tataru | Simona Senila | Chir I | George Dindelegan | Alte | OM | GS+ | GS- | Kaplan-Meier | St ...


GS_melanom_modif14.02.xls [Compatibility Mode] - Excel
Sign in
File Home Insert Page Layout Formulas Data Review View Help Acrobat Power Pivot Tell me what you want to do
Share
Cut Copy Format Painter Paste
Clipboard
Arial 10
Font
Wrap Text Merge & Center
Alignment
General
Number
Conditional Formatting Format as Table
Normal Bad Good Neutral Calculation Check Cell
Styles
Insert Delete Format
Cells
AutoSum Fill Clear Sort & Filter Find & Select
Editing
BZ1
Max_Meta
supra_DECES DECES Deces_da1_nu0 An_deces OBS Cenzura_d Cenzura_d Cenzura_l Cenzura_l CONTROL ASIG Nu_auM_a ObsNuauM OP AN_meta An_deces CONTROL ASIG ANI_PANA_META AN_pana_ An_pana_ An_pana_ Max_Dece Max_Meta
Legend RK, SS, GD, CB Rodica Cosgarea Corina Baican Al Tataru Simona Senila Chir George Dindelegan Alte OM GS+ GS- Kaplan-Meier St
100 %

62 BIAS

- Error – we must be aware of

63 EXAMPLE

- The effect of continuous training program in regulating blood pressure for people with hypertension

- Variables: group – (in program, control group), SBP, DBP (before, after)

- Age,
- weight,
- associated pathologies, etc

OUR EXAMPLE

- **The usefulness of wood crutches or aluminum crutches for people with unilateral amputation.**

- Variables: Type of crutches (wood/aluminum), Time to walk 100 m (seconds)

- Gender, Age, Leg (left/right), Sports?, etc.