

Cosmina Bondor

Hyphotesis testing II

- A** ALWAYS
- S** SEEK
- K** KNOWLEDGE

Objectives

Statistical test for frequencies:

- evaluation of the relationship between two qualitative variables
- For independent samples
 - The Chi-square test
 - The Fisher exact test
- For dependent samples
 - The Mc Nemar test

Non-parametric test

- Examples of scientific and medical articles – with explanations
- Examples of exercises for the exam

Errors in hypothesis testing

Type I error = α (alfa) – incorrect results if H_0 is true
= to reject H_0 , although it is true = false positive ($\alpha \leq 0.05$)

Type II error = β (beta)– incorrect results if H_1 is true
= to rejecting the H_1 , although it is true = false negative ($\beta \leq 0.15$)

Level of confidence, power of a test

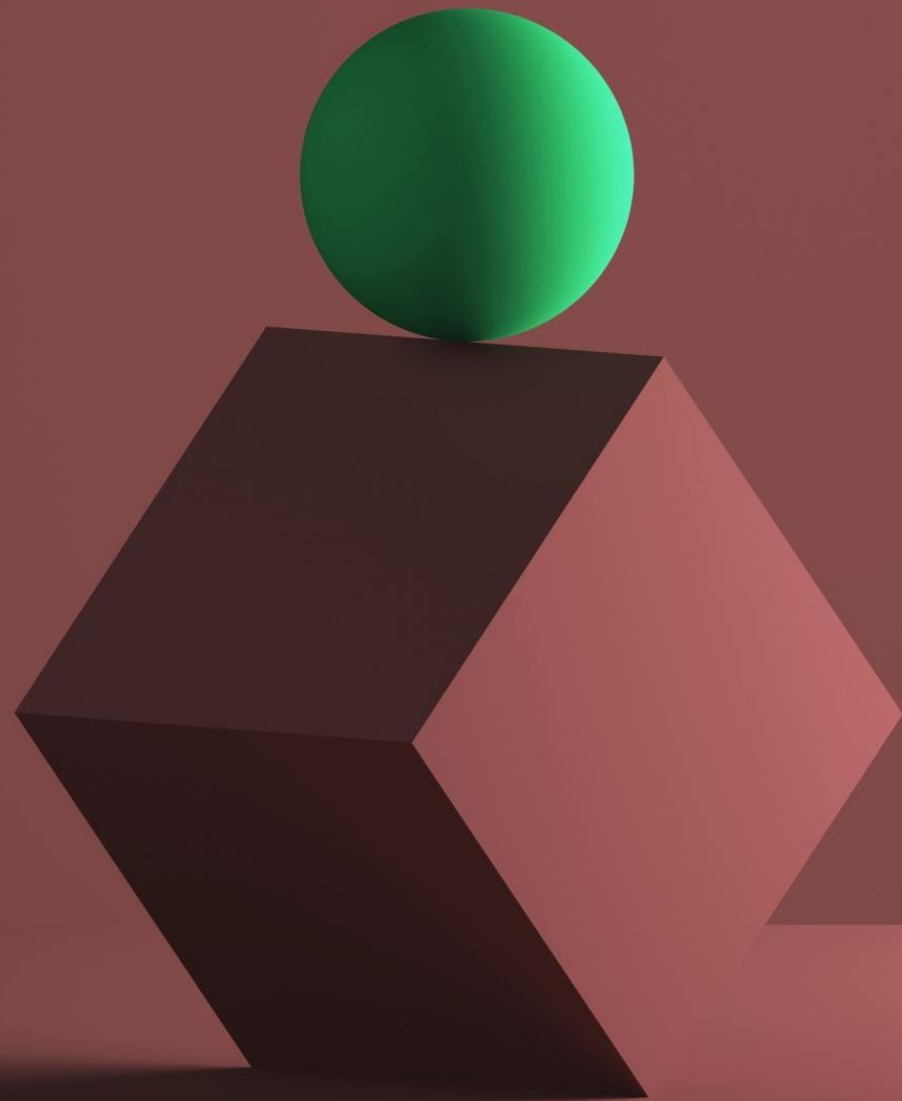
1 – α Confidence level

if $\alpha=0.05$ level of error  95% level of confidence

1 - β Test power

if $\alpha=0.15$ beta error  85% power of the test

- the power to detect a given difference of a given size, if that difference exists – the sensibility of the test
- influenced by
 - the number of individuals in a sample
 - the variation of the data
- involved in the calculation of "sample size" necessary in a study



Statistical inference – qualitative variables

Compare two frequencies

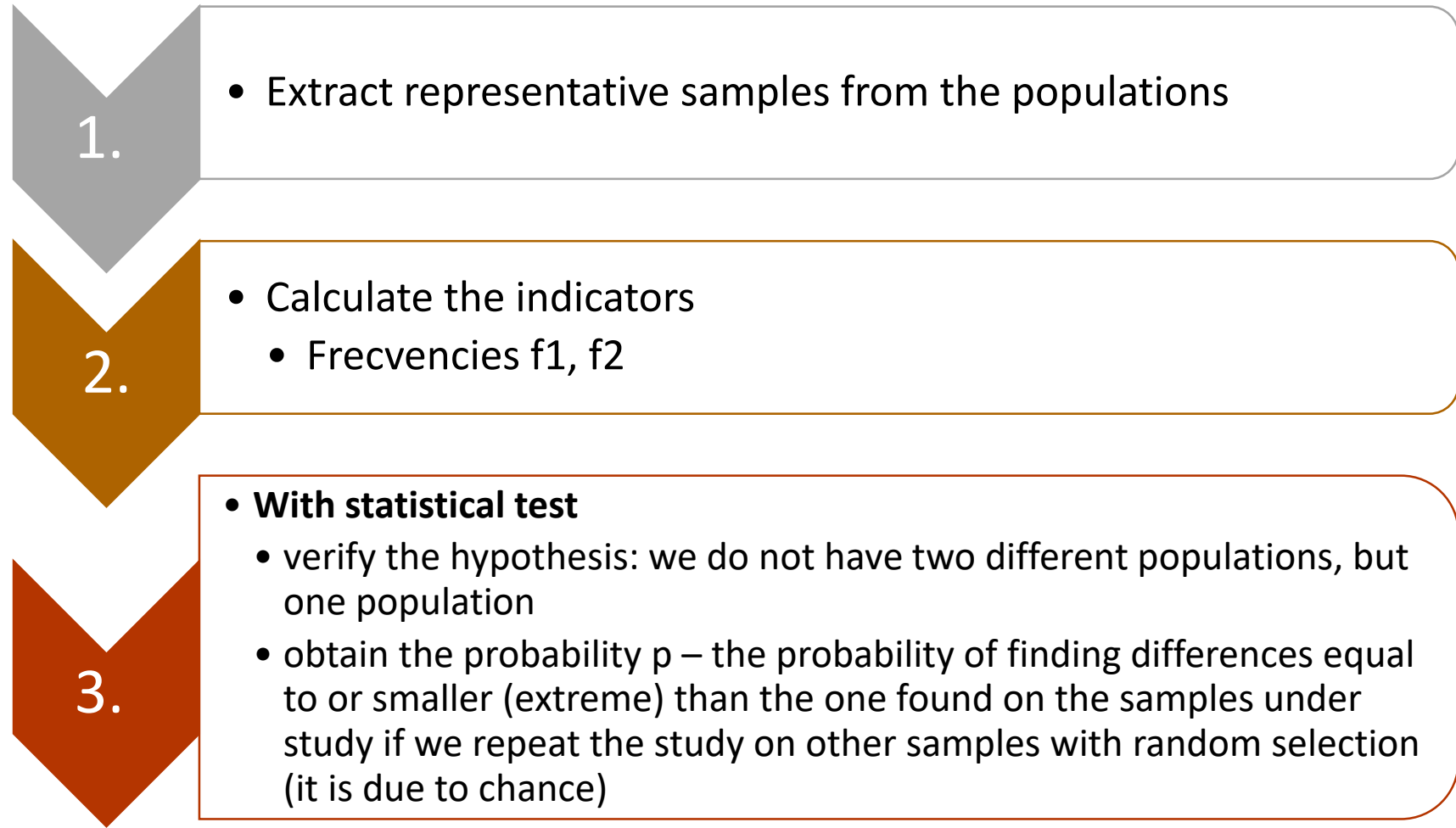
- Populations P_1, P_2, \dots
- Objective: the study of differences in the parameters of a qualitative variable X in two/more populations (categories of the grouping variable Y)

Frecvencies ϕ_1, ϕ_2, \dots

How?

- Populations P_1, P_2, \dots
- Objective: the study of differences in the parameters of a qualitative variable X in two/more populations (categories of the grouping variable Y)

Frecvencies ϕ_1, ϕ_2, \dots



How?

The hypothesis that we do not have two different populations, but the same population in terms of the frequency distribution of the variable under study

Null hypothesis H0 - assumes the denial of the objective we want to investigate

There is **no statistically significant difference** between groups in terms of frequency

There is **no statistically significant association** between 2 variables: Risk factor – disease

Alternative hypothesis H1 (denial of H0): refers to the objective we want to investigate

There is a **statistically significant difference** between groups in terms of frequency

There is a **statistically significant association** between 2 variables: Risk factor – disease

Statistical test --> we choose between the two possibilities H0 or H1

Question: In some cases, graft rejection occurs immediately after kidney transplantation. Why? Are there some factors that can be associated with graft rejection?

Hypothesis:

Obese people are more likely to experience graft rejection immediately after kidney transplantation

apply a test for frequencies
i.e. Chi-square test

Identify the variables of interest and populations

Variables

- supposed risk factor – Obesity
 - Qualitative dichotomous variable
- Disease – Graft rejection
 - Qualitative dichotomous variable

Population

- People who have had kidney transplants
- Subpopulations
 - Obese people who have had kidney transplants
 - Non-obese people who have had kidney transplants

We transform the medical question into testable statistical hypotheses

The frequency of graft rejection after kidney transplantation (up to one month) differs in

obese individuals

versus

non-obese individuals

Variables

- Risk factor – Obesity
 - Qualitative dichotomous variable
- Disease – Graft rejection
 - Qualitative dichotomous variable

Population

- People who have had kidney transplants
- Subpopulations
 - Obese people who have had kidney transplants
 - Non-obese people who have had kidney transplants

Calculate the sample size

The frequency of graft rejection after kidney transplantation (up to one month) differs in obese individuals versus non-obese individuals



We consider a **2% difference** between frequencies to be **of medical importance**

- Statistical significance $\alpha=5\%$
- Test power $\beta=80\%$



- Calculate the required sample size
- people in each sample
 - 1500 obese
 - 1500 non-obese

Conducting the study

- We randomly select from among those who are going to have kidney transplants
 - 1500 obese people
 - 1500 non-obese people
- We track the occurrence of graft rejection over a period of 1 month after the transplant

Data collection



Observed contingency table

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000

Frecvencies in the samples

Data collection



Observed contingency table

	graft rejection +	graft rejection -	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000

Frecvencies in the samples

	(%) graft rejection +
Obese ⁺ (%)	= $50/1500*100= 3.33$
Obese ⁻ (%)	= $62/1500*100= 4.13$

Sample difference: We found a difference between the frequencies of graft rejection between the obese group and non-obese group $3.33\% - 4.13\% = -0.8\%$
we want to found if this difference remain or will be higher if we repeat the study

Formulating hypotheses

Null hypothesis H0

- There is **no statistically significant difference** between obese and non-obese patients in the **frequency** of graft rejection after (up to one month) kidney transplantation
- There is **no statistically significant association** between obesity and graft rejection after (up to one month) kidney transplantation

Alternative hypothesis H1

- There is a **statistically significant difference** between obese and non-obese individuals in the **frequency** of graft rejection after (up to one month) kidney transplantation
-
- There is a **statistically significant association** between obesity and graft rejection after (up to one month) kidney transplantation

Formulating hypotheses

Null hypothesis H0

- There is **no statistically significant difference** between obese and non-obese patients in the **frequency** of graft rejection after (up to one month) kidney transplantation

both are good

- There is **no statistically significant association** between obesity and graft rejection after (up to one month) kidney transplantation

Alternative hypothesis H1

- There is a **statistically significant difference** between obese and non-obese individuals in the **frequency** of graft rejection after (up to one month) kidney transplantation

both are good

- There is a **statistically significant association** between obesity and graft rejection after (up to one month) kidney transplantation

Observed frequency table

	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

- We assume by absurdity that the null hypothesis is true
- We calculate the theoretical table in which obesity is not a risk factor

Theoretical frequency table

	+	-	Total
+	$\frac{(a + c) * (a + b)}{n}$	$\frac{(b + d) * (a + b)}{n}$	a+b
-	$\frac{(a + c) * (c + d)}{n}$	$\frac{(b + d) * (c + d)}{n}$	c+d
Total	a+c	b+d	n

Calculation of the theoretical table (null)

- We assume by absurdity that the null hypothesis is true
- We calculate the theoretical table in which obesity is not a risk factor

	graft rejection +	graft rejection -	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000

	graft rejection +	graft rejection -	Total
Obese ⁺	$=(112 \cdot 1500) / 3000 = 56$	$=(2888 \cdot 1500) / 3000 = 1444$	1500
Obese ⁻	$=(112 \cdot 1500) / 3000 = 56$	$=(2888 \cdot 1500) / 3000 = 1444$	1500
Total	112	2888	3000

Calculation of the theoretical table (null)

- Obesity is not a risk factor in this table
- There is no difference between the frequencies graft rejection between obese and non-obese $3.73\% - 3.73\% = 0$

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	56	1444	1500
Obese ⁻	56	1444	1500
Total	112	2888	3000

	graft rejection ⁺ (%)
Obese ⁺ (%)	=56/1500=3.73
Obese ⁻ (%)	=56/1500=3.73

Observed frequency table

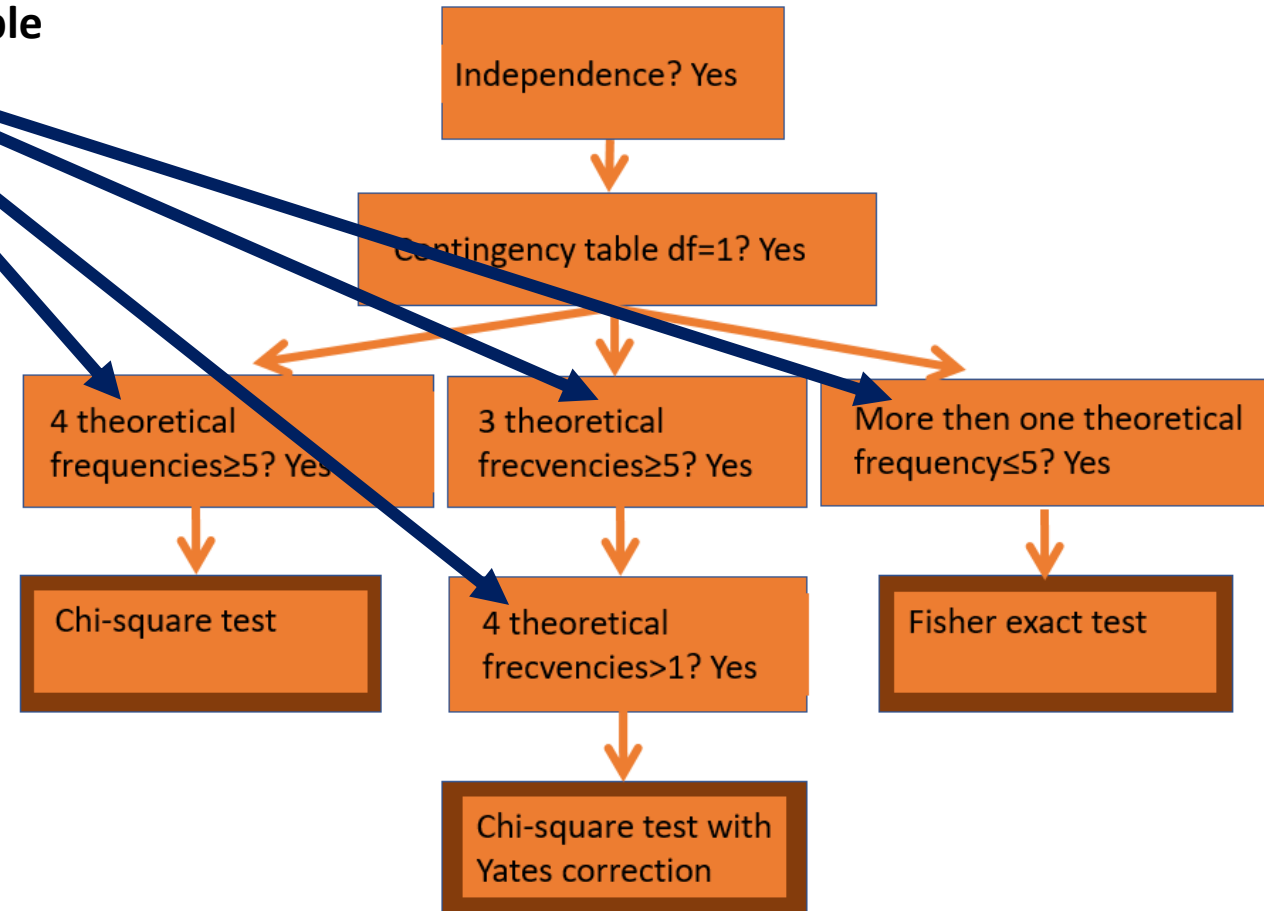
	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

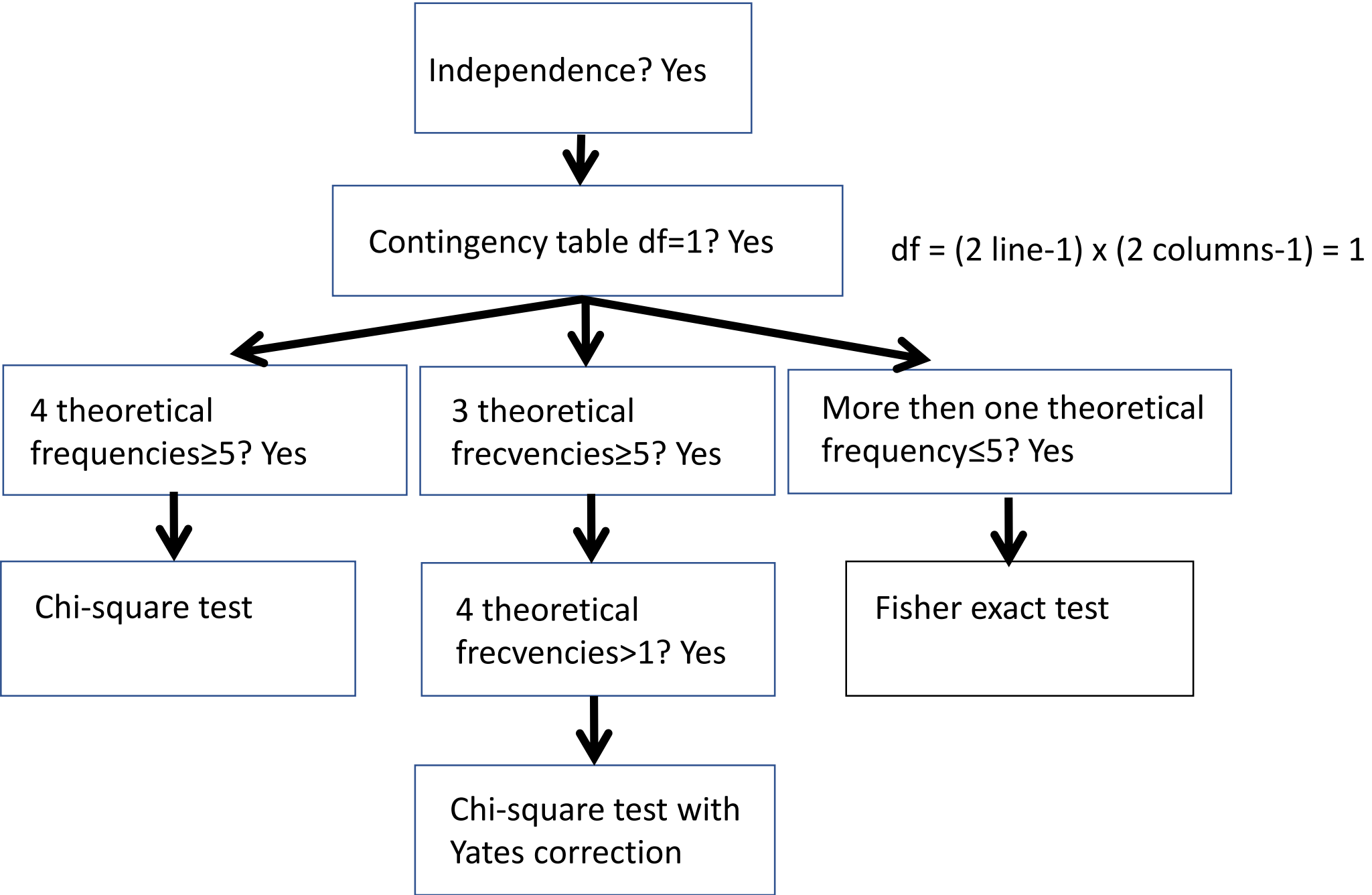
Choosing the test

Theoretical frequency table

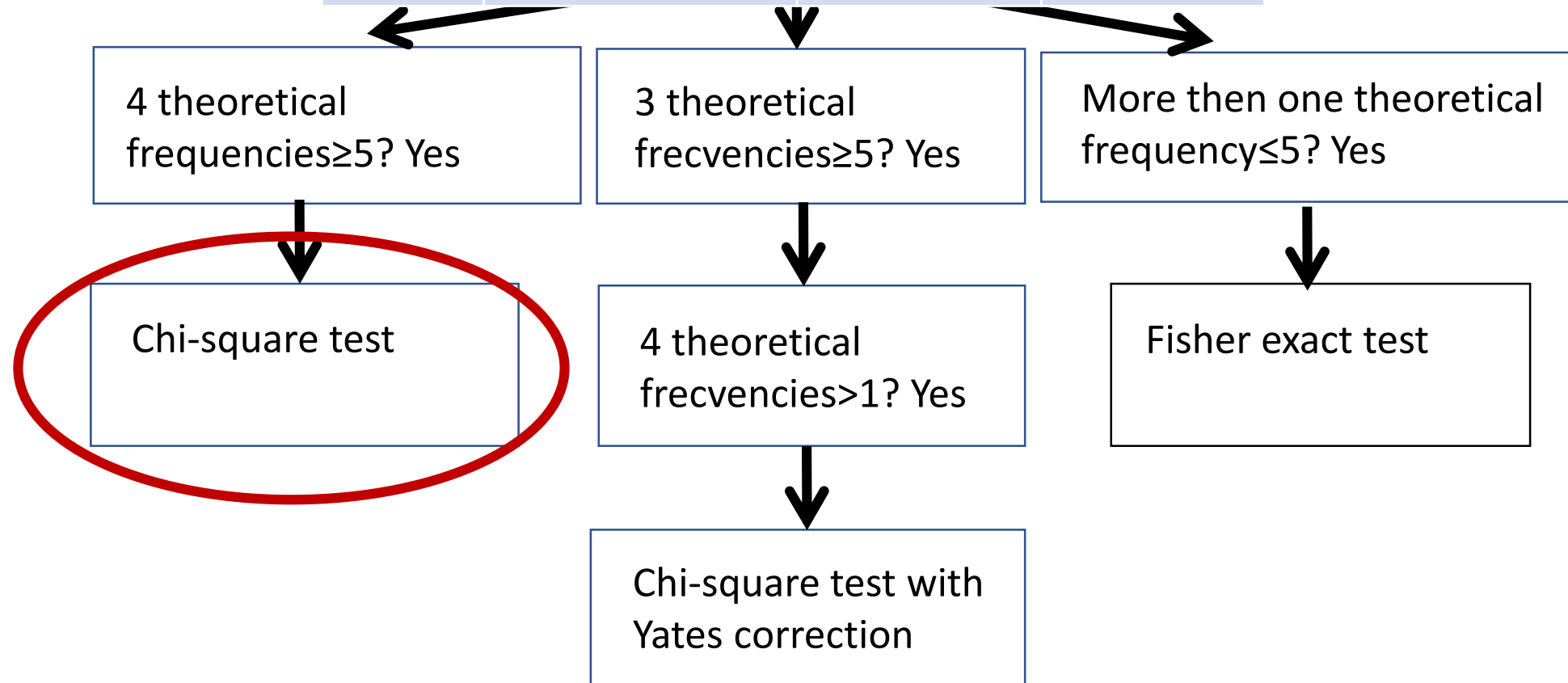
	+	-	Total
+	$\frac{(a+c) * (a+b)}{n}$	$\frac{(b+d) * (a+b)}{n}$	a+b
-	$\frac{(a+c) * (c+d)}{n}$	$\frac{(b+d) * (c+d)}{n}$	c+d
Total	a+c	b+d	n

$$df = (2 \text{ line}-1) \times (2 \text{ columns}-1) = 1$$





	graft rejection +	graft rejection -	Total
Obese ⁺	56	1444	1500
Obese ⁻	56	1444	1500
Total	112	2888	3000



- Observed contingency table

	graft rejection +	graft rejection -	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000



- Theoretic (Null) contingency table

	graft rejection +	graft rejection -	Total
Obese ⁺	56	1444	1500
Obese ⁻	56	1444	1500
Total	112	2888	3000

We calculate the **difference** between the tables

- If we find a small difference (below the critical threshold), then obesity is not a risk factor
- If we find a large difference (above the critical threshold), then obesity is a risk factor

Null hypothesis: No risk factor = no difference between the observed table and the null table

- Observed contingency table

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000



- Theoretic (Null) contingency table

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	56	1444	1500
Obese ⁻	56	1444	1500
Total	112	2888	3000

We calculate the **difference** between the tables = the Chi-square test parameter

Statistical test parameter

- expresses the difference between the compared indicators
 - e.g. frequencies, observed and theoretical contingency table
- at least one is a statistic (frequency) on the sample
- Sample/samples - randomly drawn from the population
 - a random variable follows a certain probability law
 - Ex. Chi-square law etc.

Choosing the critical region

What do we decide?

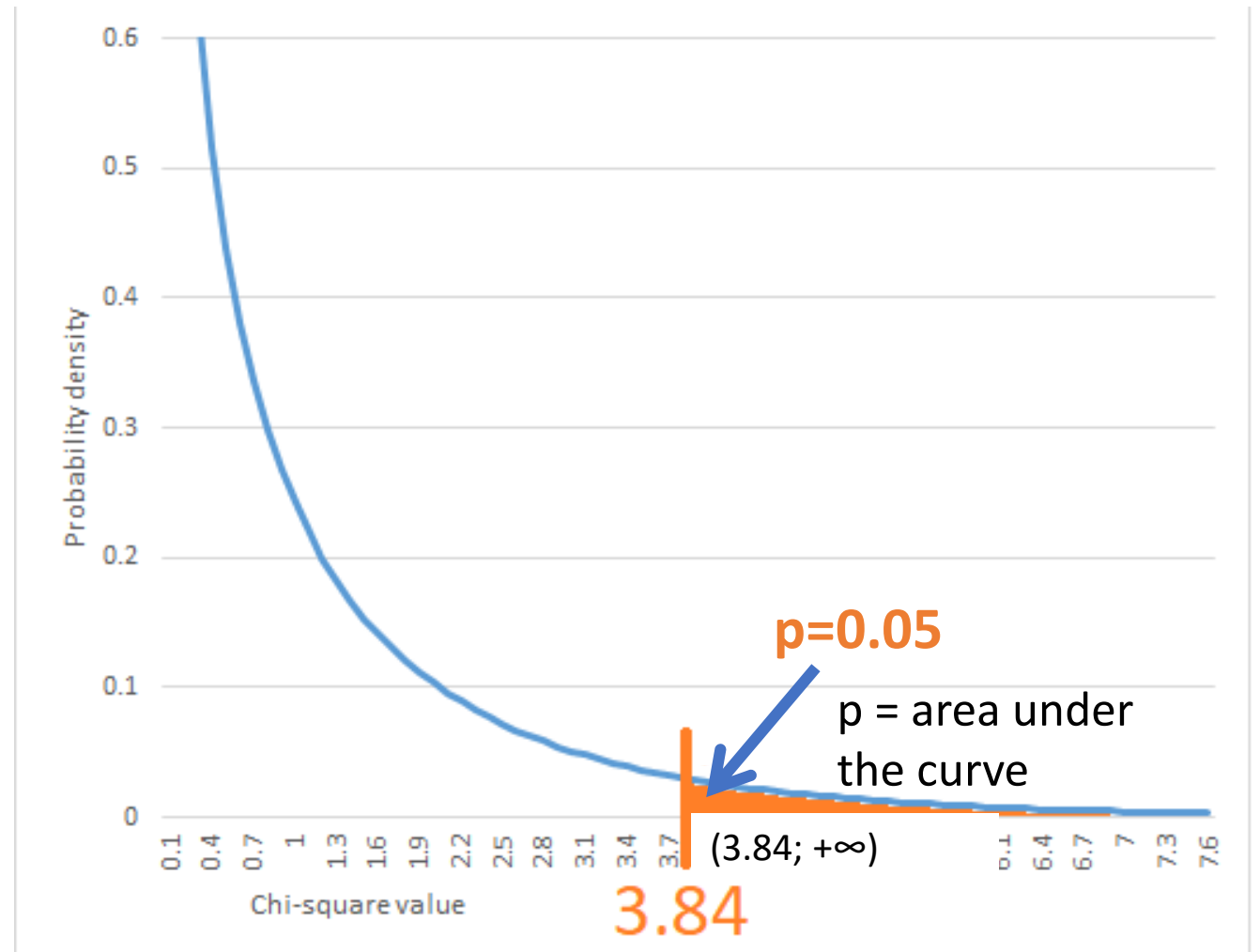
- null or alternative hypothesis,
- we decide depending on the value of the calculated parameter of the test

Choosing the size of the critical region

- depending on the size of the error risk we accept
- Significance level α = the size of the risk we are willing to assume when rejecting the null hypothesis H_0 if it is true
- Usually a significance level of 1% or 5% is chosen

Choosing the significance level and establishing the critical region

- Statistical significance $\alpha=5\%$
- Rejection region $(3.84; +\infty)$
- Acceptance area $(0; 3.84]$



test parameter follow the Chi-square distribution

- Observed contingency table

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	50	1450	1500
Obese ⁻	62	1438	1500
Total	112	2888	3000

- Theoretic contingency table

	graft rejection ⁺	graft rejection ⁻	Total
Obese ⁺	56	1444	1500
Obese ⁻	56	1444	1500
Total	112	2888	3000



parameter of the test

$$\chi^2 = \sum_{i=1}^4 \frac{(f_i^o - f_i^t)^2}{f_i^t}$$

f_i^o = observed frequency, f_i^t = theoretical frequency

- Observed contingency table

	graft rejection +	graft rejection -	Total
Obese ⁺	50	1450	
Obese ⁻	62	1438	
Total			

- Theoretic contingency table

	graft rejection +	graft rejection -	Total
Obese ⁺	56	1444	
Obese ⁻	56	1444	
Total			

$$\chi^2 = \sum_{i=1}^4 \frac{(f_i^o - f_i^t)^2}{f_i^t} = \frac{(f_1^o - f_1^t)^2}{f_1^t} + \frac{(f_2^o - f_2^t)^2}{f_2^t} + \frac{(f_3^o - f_3^t)^2}{f_3^t} + \frac{(f_4^o - f_4^t)^2}{f_4^t}$$

$$= \frac{(50 - 56)^2}{56} + \frac{(1450 - 1444)^2}{1444} + \frac{(62 - 56)^2}{56} + \frac{(1438 - 1444)^2}{1444} =$$

difference

=0.64+0.002+0.64+0.02 = 1.33

p=0.248 (from an computer application)

Test decision

A. Test decision according to the rejection region:

If χ^2 belongs to $(3.84; +\infty)$

we have enough evidence to reject H_0 , so we accept H_1

If χ^2 DOES NOT belong to $(3.84; +\infty)$

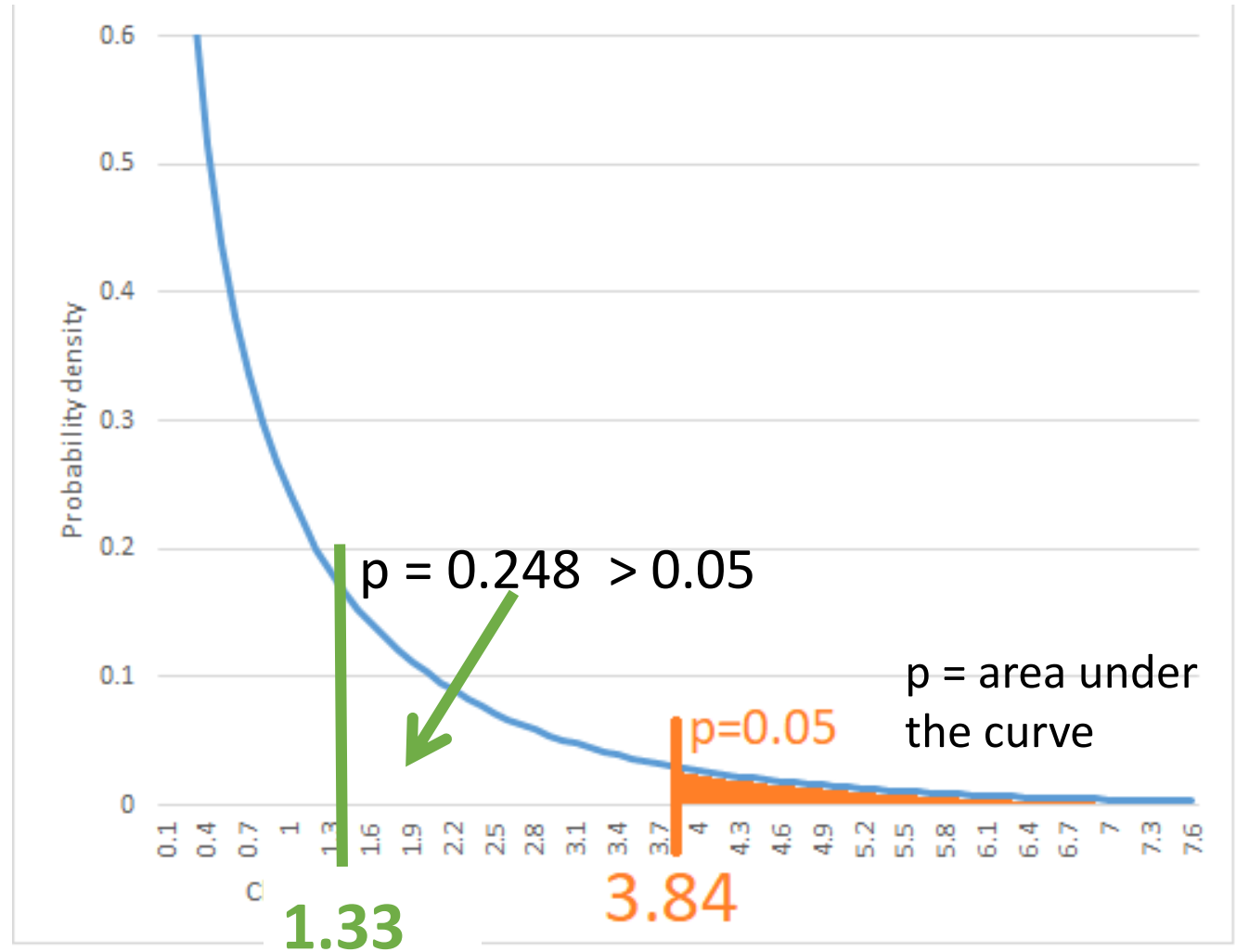
we do NOT have enough evidence to reject H_0 , so we are in favor of H_0

In our case: $\chi^2 = 1.33$ does not belong to $(3.84, +\infty)$,

we do NOT have enough evidence to reject H_0 , so we are in favor of H_0

Conclusion: There is no statistically significant association between obesity and graft rejection in the first month after kidney transplantation

$\chi^2 = 1.33$ does not belong to $(3.84, +\infty)$



B. Test decision based on p probability – the probability of finding a difference equal to or smaller than the one found if we repeat the study:

If $p < 0.05$

we have enough evidence to reject H_0 , so we accept H_1

If $p \geq 0.05$

we do NOT have enough evidence to reject H_0 , so we are in favor of H_0

In our case: $p = 0.248 > 0.05$

we do NOT have enough evidence to reject H_0 , so we are in favor of H_0

Conclusion: There is no statistically significant association between obesity and graft rejection after (up to one month) kidney transplantation, i.e. the probability of finding a difference equal to or smaller than the found one 1.33% if we repeat the study is high

Statistical decision

We were not able to show that we have two different sub-populations (obese and non-obese) in terms of graft rejection frequency, but the same population

At the population level there is no association between obesity and graft rejection in the first month after kidney transplantation

Medical conclusion

- Since we calculated the sample size and considered 2% to be a clinically significant difference
 - we can state that the 2% target was not reached,
 - differences in graft rejection rates $< 2\%$,
 - so the differences are clinically unimportant



- Obesity does not influence graft rejection in the first month after transplantation

Disadvantages of studies carried out without calculating the required sample size in advance

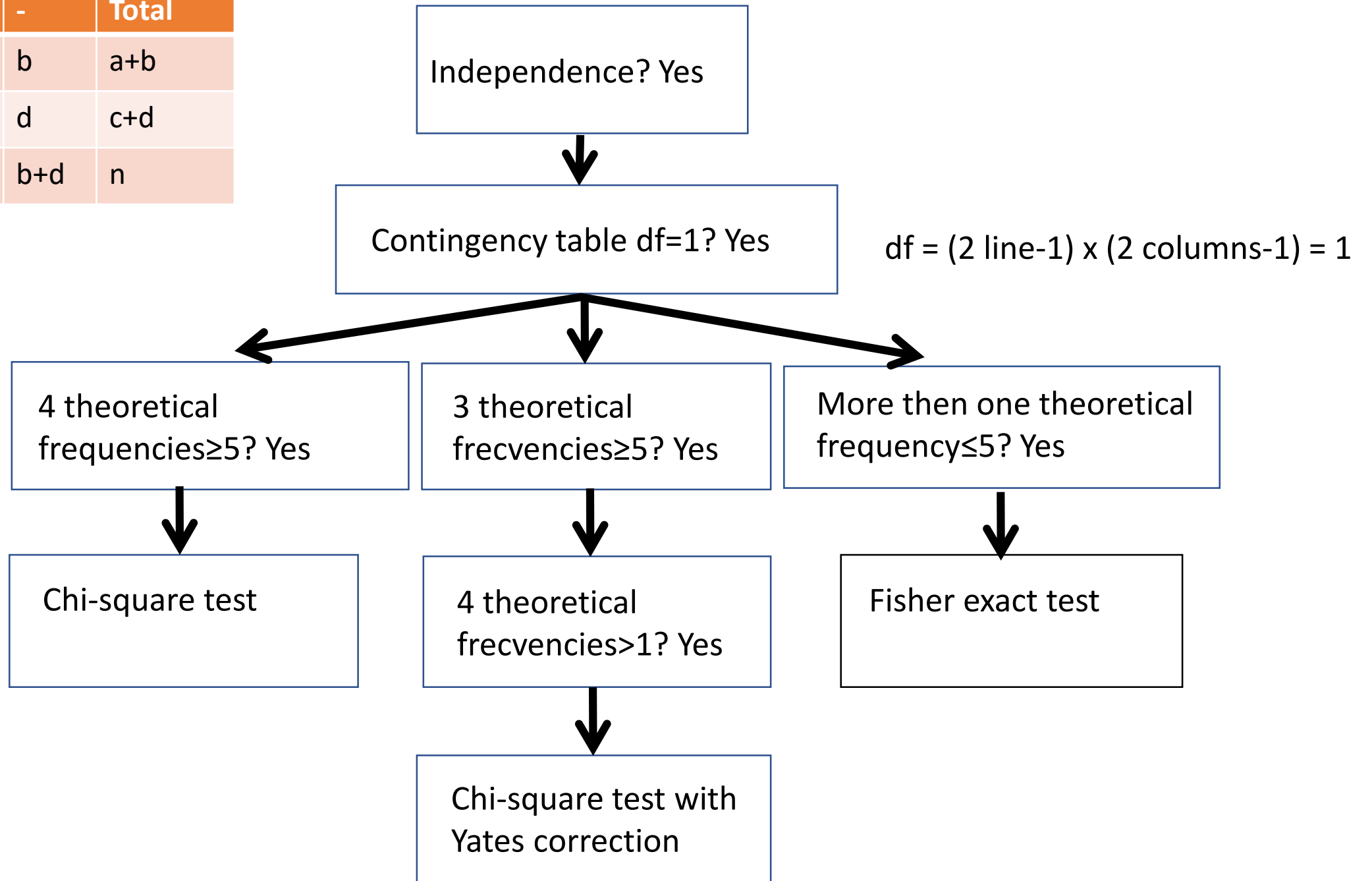
If p was significant

- We were able to demonstrate a statistically significant difference

At an insignificant p (≥ 0.05)

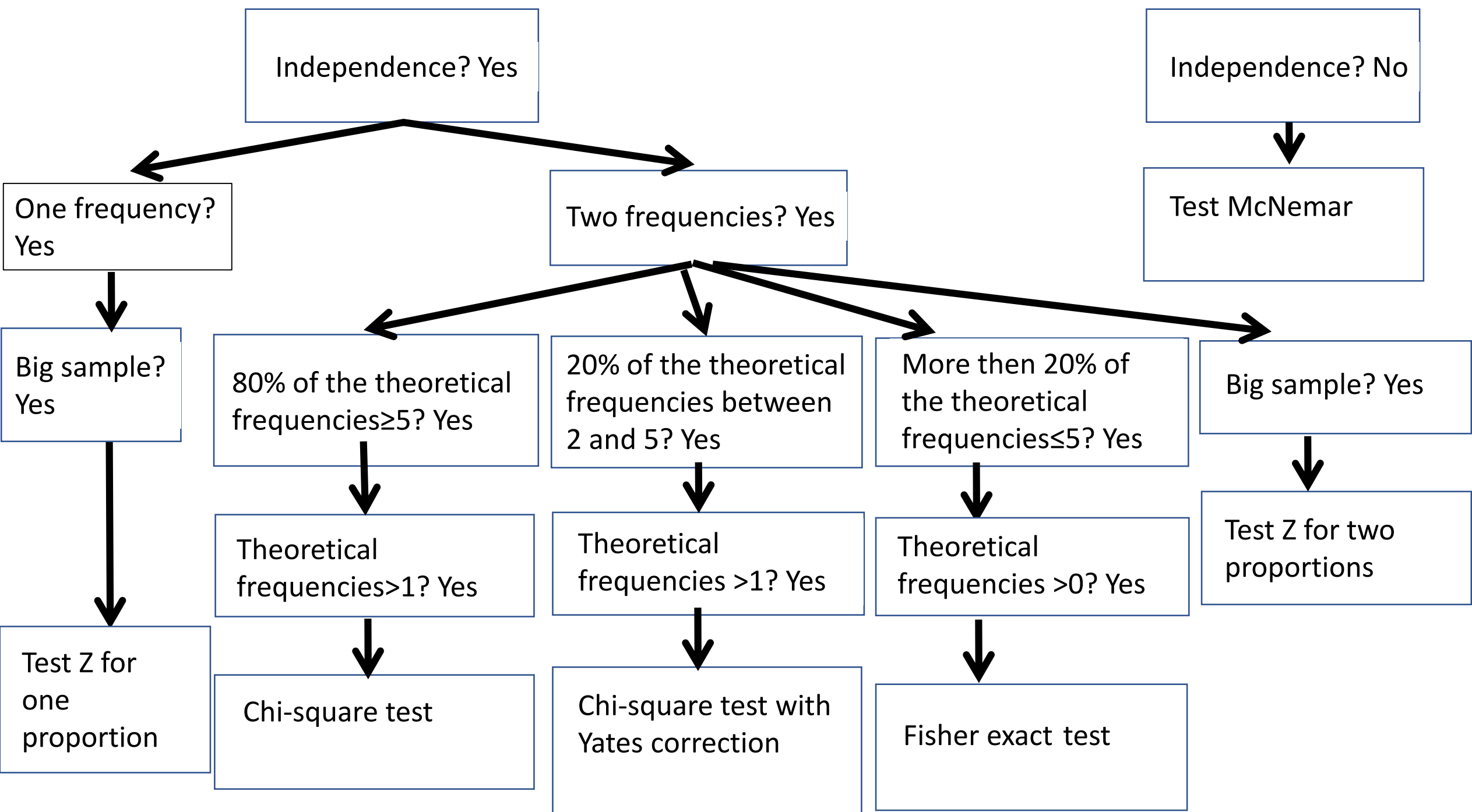
- possibly the study is not powerful enough (sample too small)
- !!! WE CANNOT DRAW THE CONCLUSION THAT THERE IS NO DIFFERENCE BETWEEN FREQUENCIES IN THE POPULATION
- If we were to draw this conclusion, it would be a TYPE I ERROR
- we drew the conclusion that there is no significant difference, although it exists

	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n



$$df = (\text{no. of column} - 1) * (\text{no. of row} - 1) > 1$$

	C1	C2	C3	...	Total
L1	a	b	c	...	a+b+c+...
L2	d	e	f	...	d+e+f+...
L3	g	h	i	...	g+h+i+...
...
Total	a+d+g+...	b+e+h+...	c+f+i+...	...	n



Assumptions

- tests for frequencies (proportions)

- two qualitative variables (or quantitative with finite discrete values)
- independent observations in each sample

Chi-square test with Yates correction

- Yates correction involves reducing by 0.5 units the difference between the observed and the theoretic frequencies within the Chi-square parameter before squaring

$$\chi^2 = \sum_{i=1}^{l \cdot c} \frac{(f_i^o - f_i^t - 0.5)^2}{f_i^t}$$

- used when >20% of the theoretical frequencies between 2 and 5
- all theoretical frequencies >1

Fisher exact test

- to test the association between two qualitative variables
- to compare independent groups (2 or >2) concerning another qualitative variable (frequencies)


Assumptions:

- two qualitative variables
- independent observations in each sample
- independent samples

It is used when we cannot use the Chi-square test.

- If at least 20% of the theoretical frequencies are less than 5

Null and alternative hypothesis – the same as for the Chi-square test

- A test that requires a lot of calculations  with computer
- p value as a result

The statistical decision

based on the p-value:

- $p \leq \alpha (=0.05) \Rightarrow$ we reject H_0 and we accept H_1
 - There is a difference / There is an association – statistically significant
- $p > \alpha (=0.05) \Rightarrow$ we cannot reject H_0
 - We can't say that there is a difference/ We can't say that there is an association – statistically significant

Mc Nemar test - for frequencies in dependent samples

- Objective: compare two frequencies studied on two dependent/paired samples

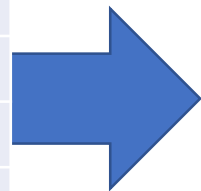
Assumptions:

- independent observations in each sample
- two dichotomous qualitative variables
- dependent/paired samples

Mc Nemar test ex.

- 200 patients suspects for facial nerve paralysis were evaluated by two neurologist to be diagnosticated in two categories with and without facial nerve paralysis. We want to see if there is an **association** between the evaluation made by medical doctors

Id patient	Observer 1	Observer 2
1	No	No
2	No	No
3	Yes	Yes
4	No	No
5	Yes	No
6	Yes	Yes
7	No	No
8	Yes	Yes
9	Yes	No
10	No	Yes



	With facial nerve paralysis Observer 1	Without facial nerve paralysis Observer 1	Total
With facial nerve paralysis Observer 2	157	2	159
Without facial nerve paralysis Observer 2	7	34	41
Total	164	36	200

	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Hypotheses:

- **Null hypothesis H0:** there is **no** association between observer 1 and observer 2 when they make the diagnostic of facial nerve paralysis
- **Alternative hypothesis H1:** there is an association between observer 1 and observer 2 when they make the diagnostic of facial nerve paralysis
- The McNemar test statistic (with continuity correction) is:
- $$\chi^2_{1ddl} = \frac{(|b-c|-0.5)^2}{b+c}$$
- follow the Chi-square distribution with 1 degree of freedom if: $b + c > 25$

3. The significance level $\alpha = 0.05$

4. The critical values and the critical region:

We find in the table of the distribution of the Chi-square the critical value X , with 1 degree of freedom

$$X_{1; 0.05} = 3.84 \Rightarrow \text{Rejection region} = [3.84, +\infty)$$

5. Calculate the value of the test statistic:

$$\begin{aligned} \chi_{1ddl}^2 &= \frac{(|b-c|-0.5)^2}{b+c} = \\ &= \frac{(|2-7|-0.5)^2}{2+7} = \frac{4.5^2}{9} = \frac{20.25}{9} = 2.25 \end{aligned}$$

	With facial nerve paralysis Observer 1	Without facial nerve paralysis Observer 1	Total
With facial nerve paralysis Observer 2	157	2	159
Without facial nerve paralysis Observer 2	7	34	41
Total	164	36	200

Statistical decision

6. The statistical decision according to the region of rejection:

- If χ^2 belongs to rejection area we can reject H_0 and accept H_1
- If χ^2 does not belong to rejection area we cannot reject H_0

- $\chi^2 = 20.32$ belongs to the region of rejection $[3.84, +\infty)$, \Rightarrow we reject H_0 with a risk of first type error ≤ 0.05

- The conclusion: we accept H_1 :
 - there is an association between observer 1 and observer 2 when they make the diagnostic of facial nerve paralysis

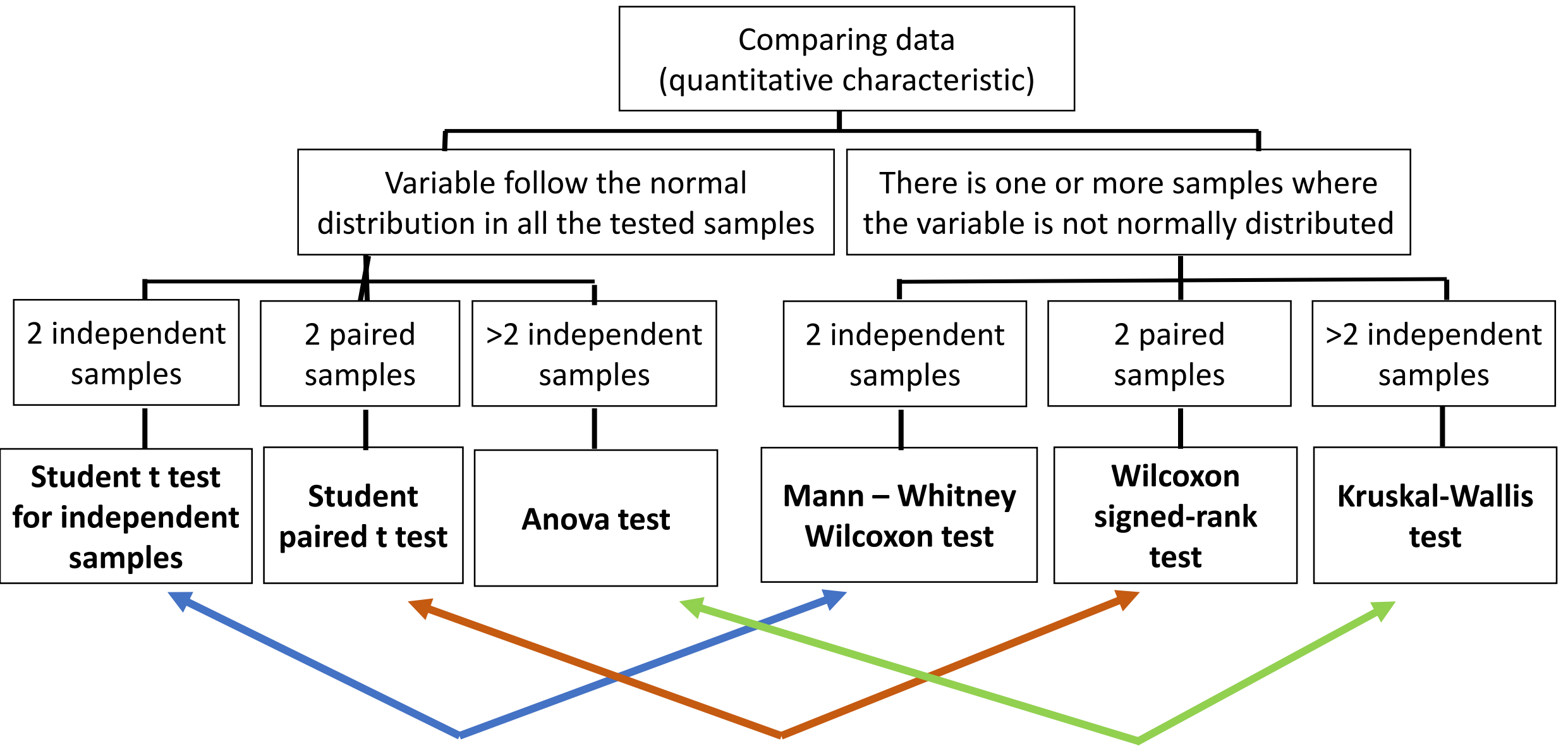
6. The statistical decision according to the p-value:

- If $p \leq \alpha (=0.05) \Rightarrow$ we reject H_0 and we accept H_1
 - There is a difference/ There is a relationship – statistically significant
- If $p > \alpha (=0.05) \Rightarrow$ we cannot reject H_0
 - We cannot say that there is a difference/ We cannot say that there is a relationship – statistically significant
- In the example: $p < 0.001 < 0.05$ we can reject H_0
 - with a risk of first type error $\leq 0.05 \Rightarrow$ we reject H_0 .

Conclusion: We accept H_1 :

- there is an association between observer 1 and observer 2 when they make the diagnostic of facial nerve paralysis

Nonparametric test



Nonparametric tests

- used as an alternative to parametric tests
 - when the data do not meet the requirements
 - e.g. do not follow the normal distribution
 - e.g. variances are not equal when applying the Anova test
- It is not necessary to use these tests
 - The data can be transformed so that they comply with the conditions
 - do not comply with the normal distribution → we apply logarithm
 - medical data will be presented using the geometric mean
 - disadvantage: Logarithmic data become difficult to interpret
 - e.g. SBP = 140 mmHg becomes $\log\text{SBP} = 4.94$
 - the data do not have equal variances → samples with an equal number of patients can be introduced into the study → the condition that the variances are equal is no longer necessary when using the Anova test

Nonparametric tests

ADVANTAGES

- assumptions are more general
- used in comparing ordinal qualitative data
- can be used in case of data with extreme values
- easy to interpret the same as the result of a parametric test

DESADVANTAGES

- are not reliable
 - the same can be said when we compare a study on a small sample study with a study on a large sample

The principle of non-parametric tests

- transforming data into their ranks

The objective of the tests

- to compare the distributions of a quantitative variable that does not meet the requirements of other parametric tests
- to compare the distributions of a qualitative ordinal variable that cannot be tested with other parametric tests
- samples can be:
 - independent
 - dependent
 - one
 - two
 - multiple

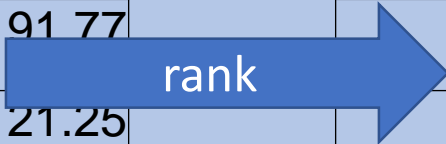
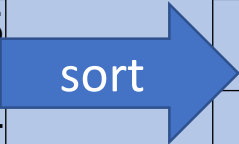
Ranks

- rank = the ordinal number of a value after sorting (ordering) the values of a variable
- we will replace the values of the variable with their rank in the order

Why?

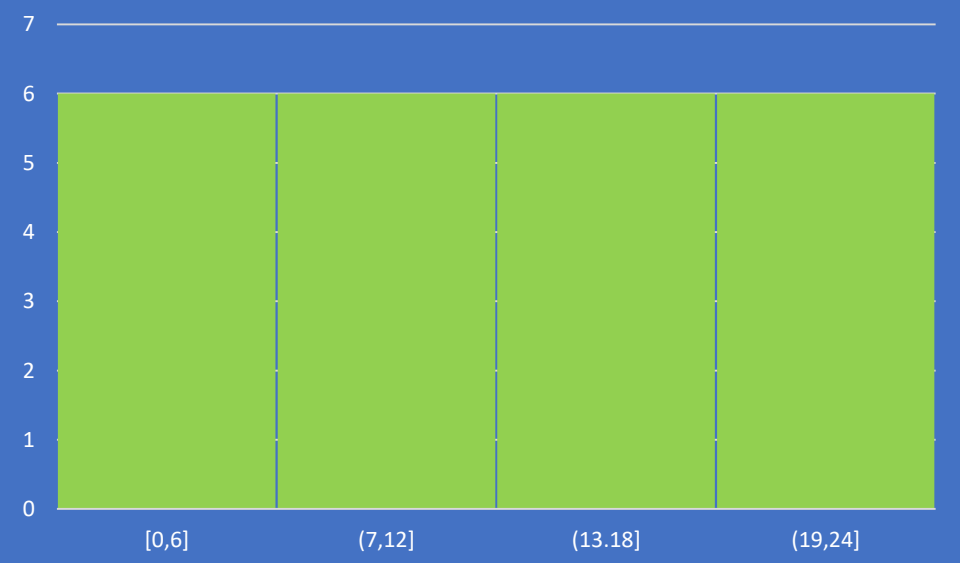
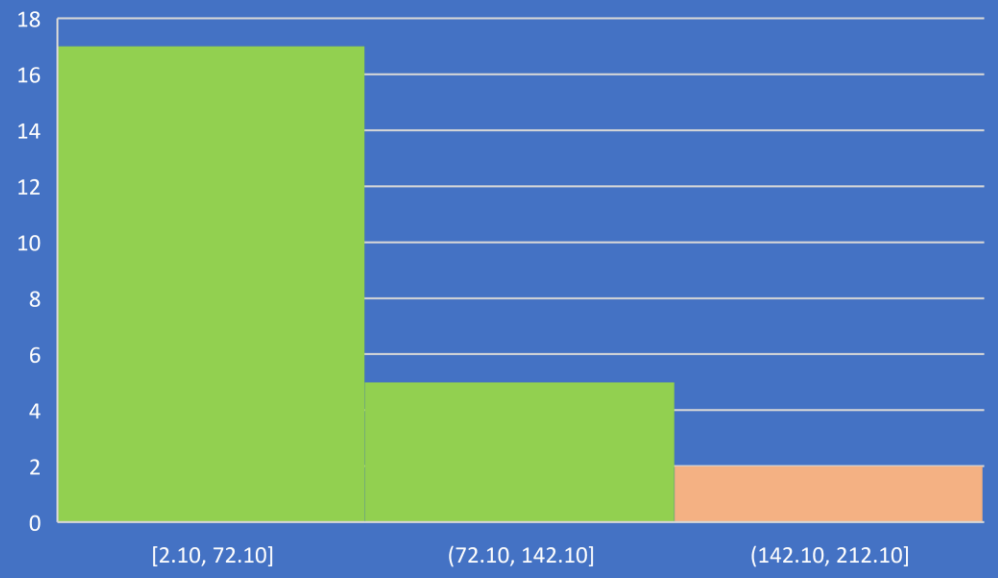
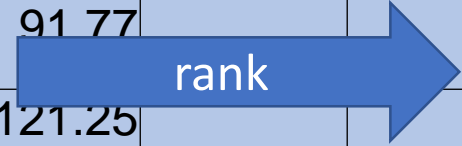
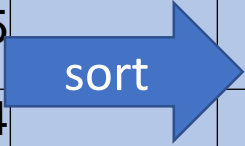
- the distribution become
 - symmetric
 - with no extreme values

C reactive protein		C reactive protein		Rank CRP	
60.40	38.38	2.10	38.38	1	13
56.13	136.77	2.24	46.09	2	14
59.88	46.09	2.27	56.13	3	15
22.06	17.87	2.28	59.88	4	16
26.00	15.44	2.43	60.40	5	17
22.40	17.59	2.45	89.31	6	18
89.31	2.45	15.44	91.77	7	19
121.25	2.24	17.59	121.25	8	20
91.77	2.43	17.87	124.75	9	21
186.57	2.28	22.06	136.77	10	22
124.75	2.10	22.40	181.73	11	23
181.73	2.27	26.00	186.57	12	24



extreme values of the variable, the assigned ranks do not contain extreme values

CRP		CRP		Rank CRP	
60.40	38.38	2.10	38.38	1	13
56.13	136.77	2.24	46.09	2	14
59.88	46.09	2.27	56.13	3	15
22.06	17.87	2.28	59.88	4	16
26.00	15.44	2.43	60.40	5	17
22.40	17.59	2.45	89.31	6	18
89.31	2.45	15.44	91.77	7	19
121.25	2.24	17.59	121.25	8	20
91.77	2.43	17.87	124.75	9	21
186.57	2.28	22.06	136.77	10	22
124.75	2.10	22.40	181.73	11	23
181.73	2.27	26.00	186.57	12	24



Equal numbers?

- for equal numbers
 - assign the same rank
 - average of their ranks
- for the next number we assign a rank by adding to the previous one the number of equal values

CRP		CRP		Rank CRP	
60	38	2	38	1,5	13
56	136	2	46	1,5	14
59	46	3	56	3	15
22	17	4	59	5	16
26	15	4	60	5	17
22	17	4	89	5	18
89	2	15	91	7	19
121	2	17	121	8,5	20
91	3	17	124	8,5	21
186	4	22	136	10,5	22
124	4	22	181	10,5	23
181	4	26	186	12	24

sort → rank

Two groups?

- Gr.1 with diabetes
- Gr.2 without diabetes
- ranks are assigned for the combined values of the two groups

Gr.1	Gr.2		Gr.1	Gr.2		Gr.1	Gr.2
6	8		2	2		2,5	2,5
5	13		2	2		2,5	2,5
9	6		4	3		6,5	5
2	7		5	4		9	6,5
6	5		6	5		12	9
2	7		6	5		12	9
8	2		8	6		18,5	12
12	2	sort	8	7	rank	18,5	15
9	5		8	7		18,5	15
8	7		9	7		21,5	15
4	3		9	8		21,5	18,5
8	4		12	13		23	24

How we compare groups?

- by calculating the average/sum of ranks
- for each group

Gr.1	Gr.2		Gr.1	Gr.2		Gr.1	Gr.2
6	8		2	2		2,5	2,5
5	13		2	2		2,5	2,5
9	6		4	3		6,5	5
2	7		5	4		9	6,5
6	5		6	5		12	9
2	7		6	5		12	9
8	2	sort	8	6	rank	18,5	12
12	2		8	7		18,5	15
9	5		8	7		18,5	15
8	7		9	7		21,5	15
4	3		9	8		21,5	18,5
8	4		12	13		23	24
6.58	5.75					14	13

Mann-Whitney U test

Nonparametric equivalent of Student's t-test for **independent samples**

Assumptions

- observations are independent within the sample
- we are comparing two independent samples
- variable of interest
 - quantitative
 - **qualitative ordered (ordinal)**

Commonly used when data do not have a normal distribution

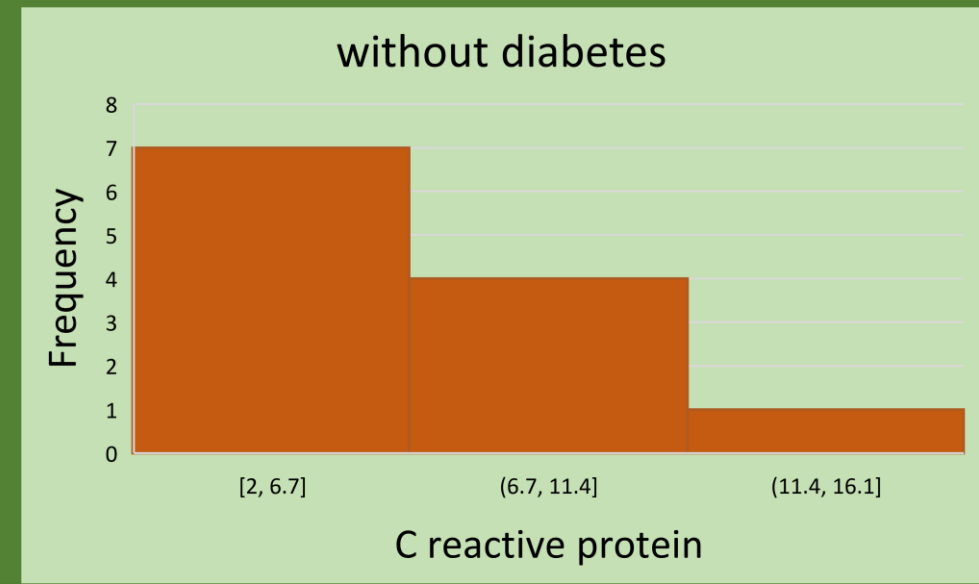
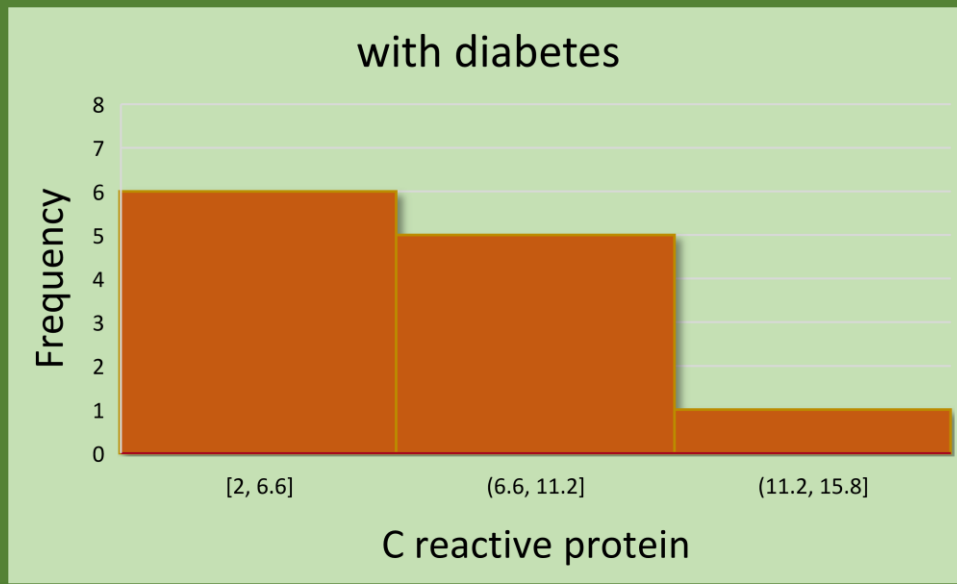
Mann-Whitney U testul

- **Null hypothesis H0:** There is no significant difference between the distributions of scores or ranks of the variable of interest in the two populations
- **Alternative hypothesis H1:** There is a significant difference between the distributions of scores or ranks of the variable of interest in the two populations
- **H1 could be formulated:**
 - there is a difference between the **mean** of the ranks associated with the values of the variable of interest in the two populations
 - there is a significant difference between the **medians** of the ranks associated with the values of the variable of interest in the two populations

Example

Objective

- C-reactive protein differs statistically significantly between patients with and without diabetes
- 12 patients were included in the study in each group
- Gr.1 with diabetes
- Gr.2 without diabetes
- Data are not normally distributed



Formulating hypotheses

- **Null hypothesis H0:** There is no significant difference between the rank distributions of CRP in those with and without diabetes
- **Alternative hypothesis H1:** There is a significant difference between the rank distributions of CRP in those with and without diabetes

Gr.1	Gr.2		Gr.1	Gr.2		Gr.1	Gr.2
6	8		2	2		2,5	2,5
5	13		2	2		2,5	2,5
9	6		4	3		6,5	5
2	7		5	4		9	6,5
6	5		6	5		12	9
2	7		6	5		12	9
8	2	→ sort →	8	6	→ rank →	18,5	12
12	2		8	7		18,5	15
9	5		8	7		18,5	15
8	7		9	7		21,5	15
4	3		9	8		21,5	18,5
8	4		12	13		23	24
6.58	5.75					14	13

Establishing the confidence level

We find the region of acceptance of the H_0

Determine the critical value U_α

- Small samples ($n_1, n_2 \leq 15$) \Rightarrow search the Mann-Whitney table for the critical value U_α
 - Acceptance region of $H_0 = [U_\alpha, +\infty)$
- Large samples ($n_1, n_2 > 15$) \Rightarrow the test statistic U follows a normal distribution $N(\mu, \sigma^2)$,
where the mean $\mu = \frac{n_1 n_2}{2}$ and the variance $\sigma^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$
- if we center and reduce we arrive at a standardized distribution $N(0, 1)$ using $Z = \frac{U - \mu}{\sigma} \Rightarrow$
 - Acceptance region of $H_0 = [-Z_\alpha, Z_\alpha]$, where $Z_\alpha =$ critical value in the table of the standardized normal law $N(0,1)$

- We take the confidence level $\alpha=0.05$

Establishing the confidence level

Finding the region of acceptance of H_0

Determine the critical value U_α

- Small samples ($n_1, n_2 \leq 15$) \Rightarrow search the Mann-Whitney table for the

critical value U_α

- Acceptance region of $H_0 = [U_\alpha, +\infty)$

- Large samples ($n_1, n_2 > 15$) \Rightarrow

- Acceptance region of $H_0 = [-Z_\alpha, Z_\alpha]$,
where Z_α = critical value in the table of
the standardized normal law $N(0,1)$

- We take the confidence level $\alpha=0.05$
- In our case $n_1 = 12, n_2 = 12$ the samples are small we look for the critical value $U_{0.05}=37$ in the Mann-Whitney table

Alpha = .05 (two-tailed)

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Establishing the confidence level

Alpha = .05 (two-tailed)

Find

Determine

- Small sample size
- the Maximum
- critical value
- Accuracy
- Large sample size
- Accuracy
- when
- the

n1 \ n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

critical value

Establishing the confidence level

Finding the region of acceptance of H_0

Determine the critical value U_α

- Small samples ($n_1, n_2 \leq 15$) \Rightarrow search the Mann-Whitney table for the critical value U_α
 - Acceptance region of $H_0 = [U_\alpha, +\infty)$
- Large samples ($n_1, n_2 > 15$) \Rightarrow
 - Acceptance region of $H_0 = [-Z_\alpha, Z_\alpha]$, where Z_α = critical value in the table of the standardized normal law $N(0,1)$

- We take the confidence level $\alpha=0.05$
- In our case $n_1 = 12, n_2 = 12$ the samples are small we look for the critical value $U_{0.05}=37$ in the Mann-Whitney table
- The acceptance region of $H_0 = [37, +\infty)$

Alpha = .05 (two-tailed)

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Calculating the U-test parameter

Calculating U as a function of n_1 , n_2 and n_i , where $i = 1$ or 2

Sum of ranks = SR

$$U_1 = n_1 n_2 + n_1 \frac{n_1 + 1}{2} - SR_1$$

$$U_2 = n_1 n_2 + n_2 \frac{n_2 + 1}{2} - SR_2$$

Test parameter $U = \min(U_1, U_2)$

$$n_1 = 12, n_2 = 12$$

Calcularea parametrului testului U

Calculating U as a function of n_1 , n_2
and n_i , where $i = 1$ or 2

Sum of ranks = SR

$$U_1 = n_1 n_2 + n_1 \frac{n_1 + 1}{2} - SR_1$$

$$U_2 = n_1 n_2 + n_2 \frac{n_2 + 1}{2} - SR_2$$

Test parameter $U = \min(U_1, U_2)$

$$n_1 = 12, n_2 = 12$$

$$SR_1 = 56, SR_2 = 104$$

$$U_1 = 12 \cdot 12 + 12 \frac{12+1}{2} - 56 = 166$$

$$U_2 = 12 \cdot 12 + 12 \frac{12+1}{2} - 104 = 118$$

$$U = \min(166, 118) = 118$$

$$U = 118$$

Statistical decision

Small samples ($n_1, n_2 \leq 15$)

$U \geq U_\alpha \rightarrow$ we cannot reject H_0

$U < U_\alpha \rightarrow$ we reject H_0 , we accept H_1

Large samples ($n_1, n_2 > 15$)

$|Z| \leq Z_\alpha \rightarrow$ we cannot reject H_0

$|Z| > Z_\alpha \rightarrow$ we reject H_0 , we accept H_1

U = 118

Decision based on probability p

$p \geq \alpha \rightarrow$ we cannot reject H_0

$p < \alpha \rightarrow$ we reject H_0 , we accept H_1

Statistical decision

Small samples ($n_1, n_2 \leq 15$)

$U \geq U_\alpha \rightarrow$ we cannot reject H_0

$U < U_\alpha \rightarrow$ we reject H_0 , we accept H_1

Large samples ($n_1, n_2 > 15$)

$|Z| \leq Z_\alpha \rightarrow$ we cannot reject H_0

$|Z| > Z_\alpha \rightarrow$ we reject H_0 , we accept H_1

Decision based on probability p

$p \geq \alpha \rightarrow$ we cannot reject H_0

$p < \alpha \rightarrow$ we reject H_0 , we accept H_1

Decision based on U

$U = 118$

$U = 118 \geq 37 \rightarrow$ we cannot reject H_0 : There is no statistically significant difference between CRP values in those with and without diabetes

Decision based on probability p

$p = 0.378$

$p = 0.378 \geq 0.05 \rightarrow$ we cannot reject H_0 : There is no statistically significant difference between CRP values in those with and without diabetes

Statistical decision

Decision based on probability p

$$p = 0.378$$

$p = 0.378 \geq 0.05 \rightarrow$ we cannot reject H_0 : There is no statistically significant difference between the rank distributions of CRP values in those with and without diabetes

Significant differences between ranks

= one of the samples is characterized by higher values of the tested characteristic, compared to the values of the other group.

it can be concluded that there are differences in the tested characteristic between the two samples.

there are no significant differences between ranks

= the values between the two samples are mixed approximately evenly

it can be concluded that there are no differences in the tested characteristic between the two samples.

Wilcoxon test

Nonparametric equivalent of Student's t-test for **dependent samples** (pairs)

Assumptions

- observations are independent within the sample
- we are comparing two dependent samples
- variable of interest
 - quantitative
 - **qualitative ordered (ordinal)**

Commonly used when data do not have a normal distribution

Wilcoxon test

- **Null hypothesis H0:** There is **no** significant difference between the distributions of scores or ranks of the variable of interest (or between the differences between paired data and zero)
- **Alternative hypothesis H1:** There is significant difference between the distributions of scores or ranks of the variable of interest (or between the differences between paired data and zero)

Example

Objective

Treatment with Glubifer (Iron +vitamins) statistically significantly improves Sideremia (Iron) values in patients with anemia

13 patients with anemia were included in the study. They were treated with Glubifer for 3 months. Sideremia was tested before (M1) and after the 3 months of treatment (M2)

Data are not normally distributed

Id patient	M1	M2	Difference d_i	Sign d_i	Rank $ d_i $
1	55	75	20	+	3
2	40	55	15	+	1
3	25	65	40	+	9
4	33	12	-21	-	4
5	44	44	0		
6	45	61	16	+	2
7	30	67	37	+	8
8	27	70	43	+	10
9	35	80	45	+	11
10	32	66	34	+	5
11	50	86	36	+	7
12	49	14	-35	-	6
13	45	95	50	+	12

differences equal to 0 are ignored

Formulating hypotheses

- **Null hypothesis H0:** There is **no** significant difference between the rank distributions of Sideremia in those with anemia before and after treatment with Glubifer
- **Alternative hypothesis H1:** There is a significant difference between the rank distributions of Sideremia in those with anemia before and after treatment with Glubifer

Id patient	M1	M2	Difference d_i	Sign d_i	Rank $ d_i $
1	55	75	20	+	3
2	40	55	15	+	1
3	25	65	40	+	9
4	33	12	-21	-	4
5	44	44	0		
6	45	61	16	+	2
7	30	67	37	+	8
8	27	70	43	+	10
9	35	80	45	+	11
10	32	66	34	+	5
11	50	86	36	+	7
12	49	14	-35	-	6
13	45	95	50	+	12

Establishing the confidence level

We find the region of acceptance of the null hypothesis

We determine the critical value W_α

Small samples ($n \leq 15$) \Rightarrow we search the Wilcoxon table for the critical value W_α

- Acceptance region of $H_0 = [W_\alpha, \frac{n(n+1)}{2} - W_\alpha]$

Large samples ($n > 15$) \Rightarrow test parameter

$$Z = \frac{T_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Z follows a standardized normal distribution $N(0, 1)$

\Rightarrow Acceptance region of $H_0 = [-Z_\alpha, Z_\alpha]$,

where Z_α = critical value in the table of the standardized normal law $N(0,1)$

- We take the confidence level $\alpha=0.05$

Establishing the confidence level

We find the region of acceptance of the null hypothesis

n	Two-Tailed Test		One-Tailed Test	
	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

- We take the confidence level $\alpha=0.05$
- In our case $n = 13$ is a small sample
- we search in the Wilcoxon distribution table
 - critical value $W_{0.05}=17$

- Acceptance area $H_0 = [17, \frac{13(13+1)}{2} - 17] = [17, 74]$

Calculating the test parameter W

Negative rank sum W^-

Positive rank sum W^+

Parameter of the test $W = \min(W^-, W^+)$

$$n = 13$$

$$W^- = 4 + 6 = 10$$

$$W^+ = 3 + 1 + 9 + 2 + 8 + 10 + 11 + 5 + 7 + 12 = 68$$

$$W = \min(10, 68) = 10$$

$$p = 0.028$$

Sign d_i	Rank $ d_i $
+	3
+	1
+	9
-	4
+	2
+	8
+	10
+	11
+	5
+	7
-	6
+	12

Statistical decision

Small sample ($n \leq 15$)

$W \in [W_\alpha, \frac{n(n+1)}{2} - W_\alpha] \rightarrow$ we fail to reject H_0

$W \notin [W_\alpha, \frac{n(n+1)}{2} - W_\alpha] \rightarrow$ reject H_0 , accept H_1

Large sample ($n > 15$)

$|Z| \leq Z_\alpha \rightarrow$ we fail to reject H_0

$|Z| > Z_\alpha \rightarrow$ reject H_0 , accept H_1

$W = 16$

- Decision based on probability p
- $p \geq \alpha \rightarrow$ we cannot reject H_0
- $p < \alpha \rightarrow$ we reject H_0 , we accept H_1

Statistical decision

Small sample ($n \leq 15$)

$W \in [W_{\alpha}, \frac{n(n+1)}{2} - W_{\alpha}] \rightarrow$ we fail to reject H_0

$W \notin [W_{\alpha}, \frac{n(n+1)}{2} - W_{\alpha}] \rightarrow$ reject H_0 , accept

Large sample ($n > 15$)

$|Z| \leq Z_{\alpha} \rightarrow$ we fail to reject H_0

$|Z| > Z_{\alpha} \rightarrow$ reject H_0 , accept H_1

- Decision based on probability p
- $p \geq \alpha \rightarrow$ we cannot reject H_0
- $p < \alpha \rightarrow$ we reject H_0 , we accept H_1

Decision based on W

$W = 10$

$W = 10 \notin [17, 74] \rightarrow$ we reject H_0 , we accept H_1 : There is a significant difference between the rank distributions of Sideremia in those with anemia before and after treatment with Glubifer

Decision based on the probability p

$p = 0.028$

$p = 0.028 < 0.05 \rightarrow$ we reject H_0 , we accept H_1 : There is a significant difference between the rank distributions of Sideremia in those with anemia before and after treatment with Glubifer

Decizia statistică

- Decision based on probability p
- $p \geq \alpha \rightarrow$ we cannot reject H_0
- $p < \alpha \rightarrow$ we reject H_0 , we accept H_1

Decision based on the probability p

$$p = 0.028$$

$p = 0.028 < 0.05 \rightarrow$ we reject H_0 , we accept H_1 :

There is a significant difference between the rank distributions of Sideremia in those with anemia before and after treatment with Glubifer

Kruskal-Wallis Test

The nonparametric equivalent of the Anova test for testing multiple means on independent samples

Assumptions

- observations are independent within the sample
- we are comparing more than two independent samples
- variable of interest
 - quantitative
 - qualitative ordered (ordinal)

Commonly used when data do not have equal variances

Kruskal-Wallis Test

- **Null hypothesis H0:** There is no significant difference between the distributions of scores or ranks of the variable of interest
- **Alternative hypothesis H1:** There is a significant difference between the distributions of scores or ranks of the variable of interest

Example

- Objective
- The depression score differs statistically significantly between the various age categories: 18-35 years, 36 – 65 years and 66 – 95 years
- 3 samples of 12,10, respectively 13 patients were included in the study
- Gr.1 $n_1 = 12$ aged between 18 - 35 years
- Gr.2 $n_2 = 10$ aged between 36 - 65 years
- Gr.3 $n_3 = 13$ aged between 66 - 95 years
- The data belong to an ordinal qualitative variable

Statistical test steps

- **Null hypothesis H0:** There is no significant difference between the distributions of depression scores in the three age groups
- **Alternative hypothesis H1:** There is a significant difference between the distributions of depression scores in the three age groups
- we take the confidence level $\alpha=0.05$

- Calculating the probability p

Decision based on the probability p

$p \geq \alpha \rightarrow$ we cannot reject H0

$p < \alpha \rightarrow$ we reject H0, we accept H₁

Influencing Medical Students' Knowledge and Attitudes Related to Disability: A Randomized Controlled Trial

Khader A Almhdawi ¹

Affiliations + expand

PMID: 33657029 DOI: [10.1097/PHM.0000000000001740](https://doi.org/10.1097/PHM.0000000000001740)

Abstract

Purposes: The aims of the study were to evaluate and to compare the efficacy of an online and a traditionally delivered undergraduate elective course in improving medical students' disability-related knowledge and attitudes.

Methods: A randomized controlled design was implemented. Participants were medical students randomly assigned into summer 8-wk disability and the society (OT100) online course, OT100 traditional course, or wellness and lifestyles (PT100) control. Demographics were collected in addition to participants' level disability knowledge and attitudes toward people with disability scale scores (1 wk before the semester and 1 wk after it). Multivariate analysis of covariance using general linear model was conducted to evaluate groups' differences in main outcome measures.

Results: In total, 198 undergraduate medical students successfully completed the study conditions: OT100 online (n = 74), OT100 traditional (n = 59), and PT100 (n = 65). OT100 groups (traditional and online) change scores were statistical similar for disability knowledge (P = 0.966) and attitudes (P = 0.705) but significantly better (P < 0.001) than the control group.

Conclusions: OT100 course delivered traditionally or online seemed effective in improving medical students' disability knowledge and attitudes toward people with disability. More studies are needed to create effective methods improving healthcare professionals' disability-related knowledge and attitudes.

- online courses
- traditional courses
- control group without courses

Influencing Medical Students' Knowledge and Attitudes Related to Disability: A Randomized Controlled Trial

Results: In total, 198 undergraduate medical students successfully completed the study conditions: OT100 online (n = 74), OT100 traditional (n = 59), and PT100 (n = 65). OT100 groups (traditional and online) change scores were statistical similar for disability knowledge (P = 0.966) and attitudes (P = 0.705) but significantly better (P < 0.001) than the control group.

traditionally delivered undergraduate elective course in improving medical students' disability-related knowledge and attitudes.

Methods: A randomized controlled design was implemented. Participants were medical students randomly assigned into summer 8-wk disability and the society (OT100) online course, OT100 traditional course, or wellness and lifestyles (PT100) control. Demographics were collected in addition to participants' level disability knowledge and attitudes toward people with disability scale scores (1 wk before the semester and 1 wk after it). Multivariate analysis of covariance using general linear model was conducted to evaluate groups' differences in main outcome measures.

Results: In total, 198 undergraduate medical students successfully completed the study conditions: OT100 online (n = 74), OT100 traditional (n = 59), and PT100 (n = 65). OT100 groups (traditional and online) change scores were statistical similar for disability knowledge (P = 0.966) and attitudes (P = 0.705) but significantly better (P < 0.001) than the control group.

Conclusions: OT100 course delivered traditionally or online seemed effective in improving medical students' disability knowledge and attitudes toward people with disability. More studies are needed to create effective methods improving healthcare professionals' disability-related knowledge and attitudes.

- p=0.966 – no difference at the population level between traditional and online courses
- p<0.001 – there was difference at the population level between the control group and the other groups

Evaluation of the utility of teaching joint relocations using cadaveric specimens

John Au ¹, Edward Palmer ², Ian Johnson ³, Mellick Chehade ⁴

Affiliations + expand

PMID: 29554966 PMCID: PMC5859743 DOI: 10.1186/s12909-018-1151-0

Abstract

Background: Like other procedural skills, the ability to relocate a joint is an important aspect of junior doctor education. Changes in the approach to teaching and learning from the traditional apprenticeship-style model have made the teaching of practical skills more difficult logistically. Workshops utilising cadaveric specimens offer a solution to this problem.

Methods: One hundred forty-six fourth year medical students were randomly divided into 5 groups. Each group received a different teaching intervention based on ankle, patella and hip relocation. The interventions consisted of online learning modules, instructional cards and workshops using skeleton models and cadaveric dislocation models. Following the intervention students were given a test containing multiple choice and true/false style questions. A 13-item 5-point Likert scale questionnaire was also delivered before and after the intervention. The data was analysed using one-way analysis of variance (ANOVA) and the Bonferroni post-hoc test.

was also delivered before and after the intervention. The data was analysed using one-way analysis of variance (ANOVA) and the Bonferroni post-hoc test.

There was no significant difference in test scores between groups exposed to cadaveric instruction on the relocation of one-, two- or all three- joints, nor any significant difference between all the cadaveric dislocation groups and the group receiving instruction on the skeleton model.

Conclusion: The results of the present study suggest that workshops utilising cadaveric dislocation models are effective in teaching joint relocation. In addition, the finding that lower fidelity models may be of equal utility may provide institutions with flexibility of delivery needed to meet financial and resource constraints.

- 146 students divided in 5 groups
- instructions
 - on skeleton models
 - on cadaveric specimens

Results: Compared to the instructional cards group, the other 4 groups showed a 10.8-19.2% improvement in total test score ($p < 0.01$) and an 18.4-25.3% improvement in self-reported understanding and confidence in performing joint relocations ($P < 0.01$). There was no significant difference in total test scores between groups exposed to cadaveric instruction on the relocation of one-, two- or all three- joints, nor any significant difference between all the cadaveric dislocation groups and the group receiving instruction on the skeleton model.

Workshops utilising cadaveric specimens offer a solution to this problem.

Methods: One hundred forty-six fourth year medical students were randomly divided into 5 groups. Each group received a different teaching intervention based on ankle, patella and hip relocation. The interventions consisted of online learning modules, instructional cards and workshops using skeleton models and cadaveric dislocation models. Following the intervention students were given a test containing multiple choice and true/false style questions. A 13-item 5-point Likert scale questionnaire was also delivered before and after the intervention. The data was analysed using one-way analysis of variance (ANOVA) and the Bonferroni post-hoc test.

Results: Compared to the instructional cards group, the other 4 groups showed a 10.8-19.2% improvement in total test score ($p < 0.01$) and an 18.4-25.3% improvement in self-reported understanding and confidence in performing joint relocations ($P < 0.01$). There was no significant difference in total test scores between groups exposed to cadaveric instruction on the relocation of one-, two- or all three- joints, nor any significant difference between all the cadaveric dislocation groups and the group receiving instruction on the skeleton model.

Conclusion: The results of the present study suggest that workshops utilising cadaveric dislocation models are effective in teaching joint relocation. In addition, the finding that lower fidelity models may be of equal utility may provide institutions with flexibility of delivery needed to meet financial and resource constraints.

- compare to skeleton models, those who learnt on the cadaveric specimens show an improvement in total test score and also on self reported confidence.

Thank you!