STATS SIMPLIFIED

A GUIDE FOR BUSY CLINICIANS TRYING TO UNDERSTAND RESEARCH!

Part 3 of 6: ★ Stat Significance

1

There are 6 SECTIONS TO THIS SERIES

1. Basics Maths
   - Fractions
   - Percentages (%, ‰, ‱)
   - Decimals

   Section 1

2. Averages and Distribution
   - Mean vs Median
   - Standard Deviation
   - 95% Confidence Intervals

   Section 2

3. Significance
   - P- value
   - Statistical significance?
   - Clinical significance?

4. Risk Comparison
   - Relative Risks - 'RR'
   - Odds Ratios – 'OR'

5. Correlations
   - Intra-class Correlation Co-efficient- 'ICC'
   - Rank Correlation – 'r-value'

6. Other
   - Intention to Treat - ITT
   - Number needed to treat - NNT

2

Let's Start!!

SIGNIFICANCE

1. Statistical Significance & p-value
2. Clinical Significance

3

A Starting Note

Unfortunately, there is no way for me to get around the fact that to understand the p-value you have to first understand the concept of the **null hypothesis** in research.

but!!! TRUST ME!

Before you do this…

4

A Starting Note

Unfortunately, there is no way for me to get around the fact that to understand the p-value you have to first understand the concept of the **null hypothesis** in research.

I am going to make this a…

5

Introduction to the Null Hypothesis

LET'S THINK…..

- Once a research trial is designed (but before actually conducting the research), researchers already know there are really only two possible outcomes:
  - Outcome 1: There **is no effect** / link found between the variables (the null hypothesis)

    eg n = 100 women with urinary incontinence → n = 50 do PFMT, n = 50 controls

    If at the end of the study the rates of urinary incontinence are exactly the same in both groups
    ie…there is no effect of doing PFMT
  - Outcome 2: There **is an effect** / link between the variables (the alternate hypothesis)

    eg n = 100 women with urinary incontinence → n = 50 do PFMT, n = 50 controls

    If at the end of the study the group doing PFMT have a significantly lower rate of urinary incontinence
    ie….. There is an effect of doing PFMT

6

The Null & Alternate Hypotheses - H_0 & H_a

- The **Null Hypothesis (H_0)** is simply the pre-study hypothesis (prediction) by the researchers that the treatment won't make any difference
- The **Alternate Hypothesis (H_a)** is the pre-study hypothesis (prediction) by the researchers that the treatment will make a difference

Examples:

1. A study aims to work out whether intra-vaginal e-stim reduces urinary incontinence
   Null Hypothesis (Ho): says that the intra-vaginal e-stim won't make any difference to urinary incontinence
   Alternate Hypothesis (Ha): says that the intra-vaginal stimulation will reduce urinary incontinence
2. A study aims to work out whether fetal birth weight impacts the chance of forceps
   Null Hypothesis (Ho): says that fetal birth weight won't have any impact on the chance of forceps
   Alternate Hypothesis (Ha): says that fetal birth weight will have an impact on the chance of forceps.

7

The Null & Alternate Hypotheses - H_0 & H_a

So there are only two hypothetical options…..

P-VALUE is about this!

The belief that there will be no effect / difference

↓

The NULL Hypothesis H_0

What is the chance this is true!

The belief that there will be an effect / difference

↓

The ALTERNATE Hypothesis H_a

8

The Null Hypothesis - H_0

To be honest, I can understand why people find the 'null-hypothesis' in research confusing.

Aren't we more interested in finding out if there is a difference as opposed to that there isn't a difference???

The first thing I will say is….

*the importance of the '**null-hypothesis**' is waaaaay easier to understand if you make a broad assumption about all **statisticians** that they are meant to be **very pessimistic people!!!***

e.g. Unlike in the legal system where you are innocent until proven guilty…..

In research…..

The assumption is that the null hypothesis is true until proven otherwise.

9

Assumption that the Null Hypothesis is true?

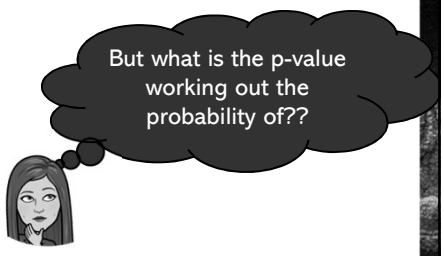
- A "Good researcher / statistician" is meant to always start by believing that the null hypothesis (ie that a treatment won't make any difference / won't be linked) is true unless there is statistical evidence otherwise.

Once you realise that basic concept about statisticians

We can now learn what the p-value is telling us!!

10

So what is the "p-value"?



- The "p" in '**p-value**' basically stands for "**probability value**"
- It is the probability that **THE NULL hypothesis is true**

NOTE: it is nothing about the alternate hypothesis, it is only about the NULL HYPOTHESIS

- in reality ….the **p-value** should probably be always written as "**Ho:p-value" =** <0.05

  - they don't put the Ho in front because Biostatisticians seem to think it is obvious that it is about the null hypothesis

- When seeing a p-value =0.57 you should say to yourself

*"The probability (ie chance) that there was actually **no difference** between the groups is 0.57"*

But a little TIP: it often becomes more understandable if you convert it to a fraction or percentage

11

P-values examples

P-value	The Likelihood that there is actually NO DIFFERENCE between the groups	Likelihood there WAS a difference b/w groups
P = 0.05	5/100 or 5%	95% probability
P = 0.64	64/100 or 64%	36% probability
P <0.001	<1/1000 or < 0.1%	99.9% probability
P = 0.02	2/100 or 2%	98% probability

Which is why we like the p-value (the probability that there wasn't a difference) to be <0.05 (ie <5%)

Note:
In research we like there to be >95% chance that there was a real difference for us to consider believing the result is true.

12

Let me show you how this works in two research statements.....

Example 1: A research paper states that

"after administering a pain relieving drug the treatment group experienced a 3/10 (SD 0.5) drop in pain, compared to a 1/10 (SD 0.3) drop in pain in the placebo group (p =0.02).

From our previous learning on mean and standard deviation:
TREATMENT GROUP
1. The average drop in pain in the treatment group was 3/10, and…..
2. Based on the Stand. Dev = 0.5: **95% of the Rx group probably** experienced a **drop between 2/10 and 4/10**

CONTROL / PLACEBO GROUP
1. The average drop in pain in the treatment group was 1/10, and
2. Based on the Stand. Dev. = 0.3: **95% of the Rx group** probably experienced a **drop between 0.4 - 1.6/10**

so what is the p-value telling us?

13

Let me show you how this works in two research statements.....

Example 1: A research paper states that

"after administering a pain relieving drug the treatment group experienced a 3/10 (SD 0.5) drop in pain, compared to a 1/10 (SD 0.3) drop in pain in the placebo group (p =0.02).

The p-value (p = 0.02) tells us that based on this sample, there is only a 2% chance that there will be no difference between the groups when treating with this drug. (ie the null hypothesis is very unlikely)

Or……. In alternate terms:

It tells us that there is a **98% chance that the difference we saw is actually true**

14

Let me show you how this works in two research statements.....

Example 2: A research paper states that

"after administering a pain relieving drug the treatment group experienced a 4/10 (SD 1.5) drop in pain on a VAS pain scale vs a 2/10 (SD 0.5) drop in the placebo group (p =0.44).

If we first look only at the MEANS…..
1. The average drop in pain in the treatment group was 4/10, and…..
2. The average drop in pain in the control group was 2/10

BUT….. What if we consider the standard deviation (the range of different results in participants)

TREATMENT GROUP:
95% of the Rx group experienced a drop somewhere between 1/10 and 7/10 (2SD either side of the mean)
CONTROL GROUP
95% of the control group experience a drop somewhere between 1/10 and 3/10 (2SD either side of the mean)

15

Let me show you how this works in two research statements.....

Example 2: A research paper states that

"after administering a pain relieving drug the treatment group experienced a 4/10 (SD 1.5) drop in pain on a VAS pain scale vs a 2/10 (SD 0.5) drop in the placebo group (p =0.44).

TREATMENT GROUP:
95% of the Rx group experienced a drop somewhere between 1/10 and 7/10 (2SD either side of the mean)
CONTROL GROUP
95% of the control group experience a drop somewhere between 1/10 and 3/10 (2SD either side of the mean)

But the p-value was 0.44

The p = 0.44 says to us that based on this data there is a 44% chance that the null hypothesis is actually true and there will be no real difference between the groups when treating with this drug.

Ie….. **We really can't be sure that the difference in the means is a true statistically significant difference.**

16

THE MAIN POINT

If the p-value is > 0.05 it means that there is a more than 5% chance that the null hypothesis is actually true and there is no difference between groups.

There is therefore *usually* not much point making decisions based on the result

IN CONTRAST

If the p-value is <0.05 then there is a less than 5% chance that there was actually no difference…..

or in other words ……

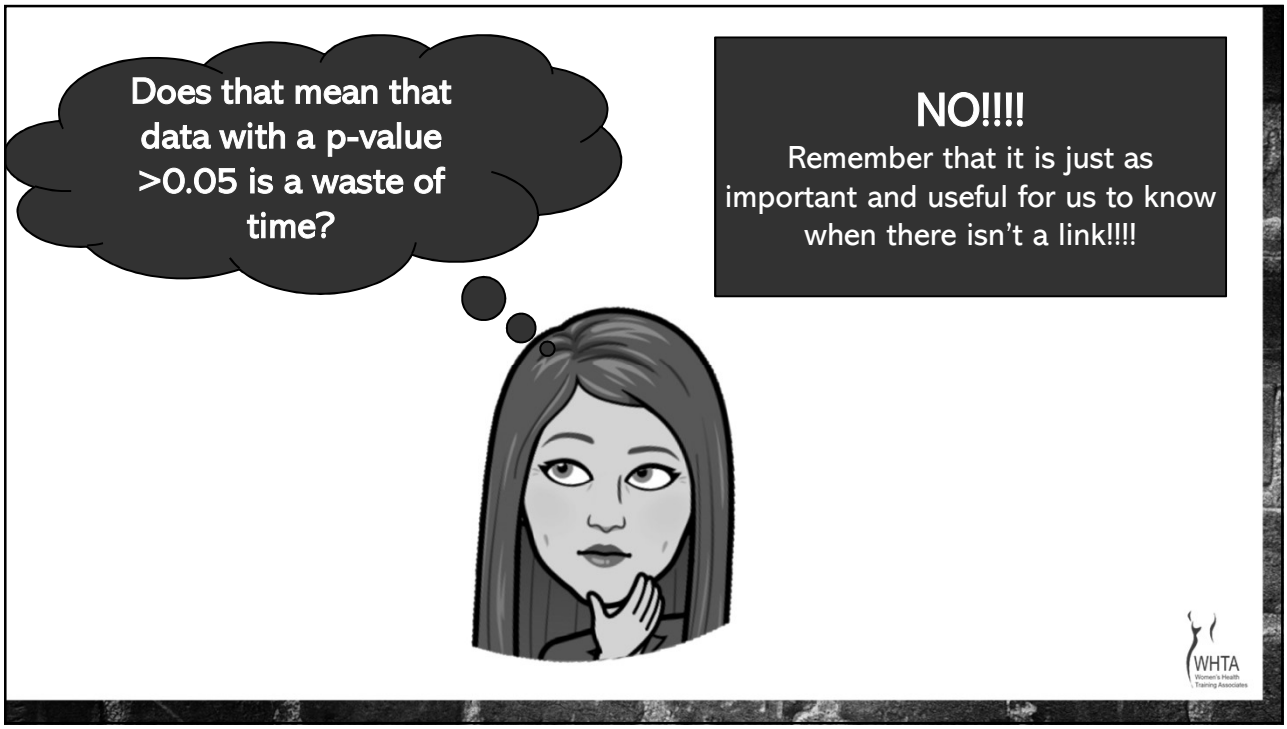
there is a more than 95% chance that the difference was real and not a fluke result of the data.

17

Let's look at some more examples

STATEMENT	WHAT IT MEANS
The PFMT group showed a mean reduction in incontinence episodes per day of 3/day compared to the control group 1/day (p<0.01)	1. The data suggests that PFMT reduced the Rx groups incontinence episodes per day by 2 more than the control group. AND!! 2. Based on the data there is a <1% chance there would be no difference between the groups in real life (99% chance there was a true difference)
The bladder retraining group showed a reduction in nocturia episodes per night of 1.2/night vs 0.7/night in the control group (p = 0.64)	1. The data suggests that bladder retraining reduced the nocturia episodes per night by 0.5/night more than the control group. BUT!!! 2. There is a 64% chance that there is actually no real difference between the groups.
Women with forceps had an avulsion rate of 36% compared to those with a NVD who had an avulsion rate of 12% (p = 0.001)	1. The data suggests that women who had forceps were three times more likely to have an avulsion than those who had a normal vaginal delivery (36% vs 12%). AND!! 2. Based on the data there is only a 1/1000 chance that this isn't true and there is actually no difference between the groups.

18



19

Example of a time where a high p-value is good to know

A research paper compares the rate of forceps in women who did and did not do PFMT during their pregnancy.

RESULTS

- Rate of Forceps deliveries in the PFMT group was 14%
- Rate of Forceps deliveries in the no-PFMT group was 11%

BUT!!! The p-value regarding the difference was p = 0.72

WHAT DOES THIS MEAN?

Although at first glance it looks like the PFMT group might have had a higher rate of forceps, further analysis of the data tells us that there is a 72% chance though that there was actually no real difference between the groups once they looked at the demographics of the sample, standard deviation etc

THIS IS WHEN WE CALL A RESULT "NOT STATISTICALLY SIGNIFICANT"

WHTA

20

THE MAIN POINT TO REMEMBER

- **If p < 0.05**: There is a less than 5% chance there is no difference between groups

  There is a 95% chance there is a difference between the groups

- **If p >0.05**: There is a more than 5% chance there is no difference between the groups

  There is too much chance there is no difference, so at the moment we don't consider there to be a proven relationship with the factor being researched.

21

1. Statistical Significance & p-value 

This is the main part done!

We just need to now briefly discuss…..

2. **Clinical Significance**

Little Note
I really think this is not considered enough in research

22

Let's Think.....

What makes a treatment not only evidenced based, but also justified to implement??

23

Evidence Based Justifiable Treatments

To be a justifiable treatment, the treatment needs to actually achieve three criteria……

Result in a Statistically significant difference in symptoms compared to not doing anything

Result in a **Clinically Significant** Difference in Symptoms

ie noticeable difference to the patient

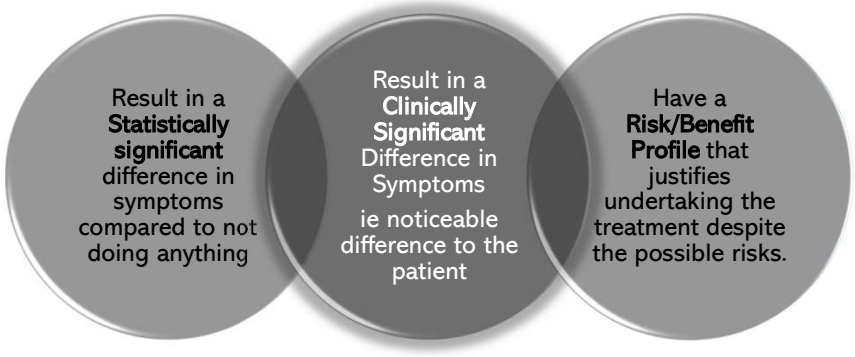
Have a **Risk/Benefit Profile** that justifies undertaking the treatment despite the possible risks.

This is what our p-value tells us…….
There needs to be a less than 5% chance that there is actually no difference between groups.

24

Evidence Based Justifiable Treatments

ie A CLINICALLY SIGNIFICANT DIFFERENCE



It can't be a difference that is simply 'statistically different',
It needs to be noticeable to the patient

25

Minimal Important Change (MIC)

- A term introduced by Jaeschke et al in 1989

PAPER TITLE

"Measurement of health status, ascertaining the minimal clinically important difference",

Controlled Clinical Trials, 1989, vol 10, 407 – 415

- The Minimal Important change (MIC) is defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient's management".

26

Minimal Important Change (MIC)

How is the Minimal Important Change defined for any outcome parameter

Usually by doing research where patients rate their change in symptoms on two scales at the same time

7-Point Global Index of Patient Improvement

1. Very much better
2. Much better
3. A little better
4. The same
5. A bit worse
6. Much worse
7. Very much worse

Scale Being Tested
eg 24hr Pad Weigh

The level of changed needed on the scale being tested….

Is correlated with what resulted in a change of at least 'a little bit better' on the patient perception of improvement

This is then defined as the 'Minimum' or 'Minimal' important change for the specified outcome measure

27

Example: Pelvic Floor Distress Inventory - 20

- The PFDI-20 is the short form version of the full length PFDI
- It is a validated questionnaire to assess the level of distress from Urinary, Bowel and POP disorders.
- Consists of a total of 20 questions each scored from 0 (no symptom) to 4 (significant bother)

Calculating the Score

Calculating the final score is a little complex and more than I want to get into here.

Simplified…. The questions are split into three sub-scales that then get multiplied resulting in:

PFDI-20 = Total Score /300

28

Research on the MIC on the PFDI-20

Barber et al 2005 (original PFDI-20 reliability study)

- Calculated the change in the PFDI-20 from baseline to 6-months after POP surgery, and compared against a global index of improvement 7-point scale rated by the patients at the same time.
- 7-Point Global Index of Improvement Scale ranged from
  - I am feeling very much worse → I am feeling very much better
- The Minimum Important Change on the PFDI-20 was determined by calculating the mean change in PFDI score of subjects indicating that they were a 'little bit better" on the global rating scale after having POP Surgery.

Barber, M.D., Walters, M.D. and Bump, R.C., 2005. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). *American journal of obstetrics and gynecology, 193*(1), pp.103-113.

29

Research on the MIC on the PFDI-20

Barber et al 2005 (original PFDI-20 reliability study)

RESULTS

Global Perception of Change	Change on PFDI-20 / 300
Worse (a little, much or very much)	+22
A little better	-45
Much Better	-73
Very much Better	-106

CONCLUSION

1. Based on a minimal important change of feeling a "little bit better", it is estimated that the minimal change on the **PFDI-20 needs to be 45points** (out of a total of 300)
2. A feeling of "much better" correlated to a change in PFDI-20 of 73 points

Barber, M.D., Walters, M.D. and Bump, R.C., 2005. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). *American journal of obstetrics and gynecology, 193*(1), pp.103-113.

30

Let's see how this looks in real life....

N = 100 women with Stage 2 and 3 uterine prolapse are randomly allocated to either

- N = 50 → undergoing treatment via PFMT
- N = 50 → control group with no treatment

Pre-Treatment both groups were comparable for age, BMI, Stage of POP, and PFDI-score

	PFMT GP	CONTROL GP	p-value
PRE-TREATMENT PFDI-20	146	143	0.64
POST-TREATMENT PFDI-20	126	139	0.04
Difference	-20	-4	0.01

This authors report that **"after PFMT, the treatment group showed a statistically significant reduction in distress (p= 0.01)"**

31

WHAT DOES THIS CONCLUSION MEAN?

	PFMT GP	CONTROL GP	p-value
PRE-TREATMENT PFDI-20	146	143	0.64
POST-TREATMENT PFDI-20	126	139	0.04
Difference	-20	-4	0.01

This authors report that **"after PFMT, the treatment group showed a statistically significant reduction in distress (p= 0.01)"**

The p = 0.01 simply means that the chance there is no real difference between groups is only 1%

BUT…… it doesn't say that this difference is significant!

The previous study showed that women needed at least a 45 point change to even perceive a "little improvement"

32

WHAT DOES THIS CONCLUSION MEAN?

	PFMT GP	CONTROL GP	p-value
PRE-TREATMENT PFDI-20	146	143	0.64
POST-TREATMENT PFDI-20	126	139	0.04
Difference	-20	-4	0.01

This authors report that **"after PFMT, the treatment group showed a statistically significant reduction in distress (p= 0.01)"**

This ultimately means that PFMT resulted in

1. A statistically significant improvement compared to control group

but

2. It would not be clinically significant (perceptible) to the patient.

33

Let's compare to one more example....

N = 100 women with SUI randomly allocated to either

- N = 50 → undergoing treatment via PFMT
- N = 50 → control group with no treatment

Pre-Treatment both groups were comparable for age, BMI & 1hr pad test at baseline

	PFMT GP	CONTROL GP	p-value
PRE-TREATMENT Pad Test (g)	16.7g	17.1g	0.37
POST-TREATMENT Pad Test (g)	2.1g	14.2g	0.001
Reduction in Pad Test (g)	-14.6	-2.9g	<0.001

This authors report that **"after PFMT, the treatment group showed a statistically and clinically significant reduction in 1hour pad test of 14.6g vs 2.9g (p<0.001)"**

34

That's it!!

SIGNIFICANCE

1. Statistical Significance & p-value
2. Clinical Significance

35

You can now say……

SIGNIFICANCE

1. Statistical Significance & p-value
2. Clinical Significance

36

What's Next??

1. Basics Maths	2. Averages and Distribution	3. Significance
Fractions	Mean vs Median	P- value
Percentages (%, ‰, ‱)	Standard Deviation	Statistical significance?
Decimals	95% Confidence Intervals	Clinical significance?
Section 1	Section 2	

4. Risk Comparison	5. Correlations	6. Other
Relative Risks - 'RR'	Intra-class Correlation Co-efficient- 'ICC'	Intention to Treat - ITT
Odds Ratios – 'OR'	Rank Correlation – 'r-value'	Number needed to treat - NNT

37

38