Medicine & the **Media**
The Challenge of Reporting on Medical Research

Limited Role of Statistics

Steven Woloshin, MD, MS
Center for Medicine in the Media
Dartmouth Institute, Lebanon, NH, USA

Il Pensiero Scientifico Editore

ProMIS

1

Why results might be wrong?

May not be valid

Biased study design or execution

Confounding: Heart attacks may be lower in supplement takers but it may have nothing to do with the supplements.

Just reflect chance

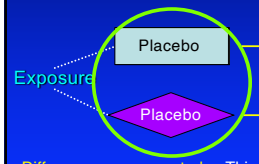
Flip a coin 10 times and get 4 heads

Flip it another 10 times and get 7 heads

Differences are expected

2

Thought experiment



Count Death

Count Death

But exactly??

Should be the same!

Differences are expected: Things vary just by chance*

Synonyms: luck of draw, random variation, sampling error

3

p-value

Is the difference bigger than what you'd expect just by chance?

The p value says how often you would expect to see a difference this big (bigger) IF IN FACT treatment really had no effect*.

Just like in the thought experiment where we tested the effect of placebo vs. placebo in identical groups of patients – in which case, the observed differences only reflect chance.

4

p-value

Is the difference bigger than what you'd expect just by chance?

The p value says how often you would expect to see a difference this big (bigger) IF IN FACT treatment really had no effect*.

p-values are calculated by:

- Assuming the drugs are the same: "null hypothesis"
- Judging if the observed data disprove this assumption

5

p-value

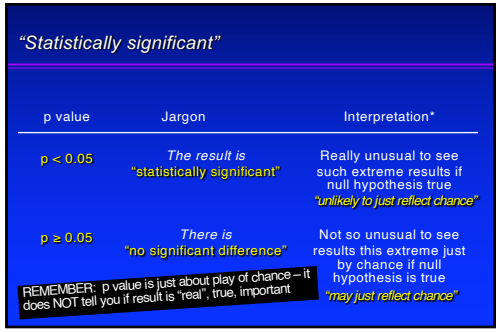
The "p" means "probability", (so it ranges from 0 to 1):
eg, p=0.03 means a 3% chance of seeing a difference this big or bigger if treatment actually had no effect

Low p value means you are less likely to see results this extreme if null hypothesis were true (ie, treatment really made no difference)

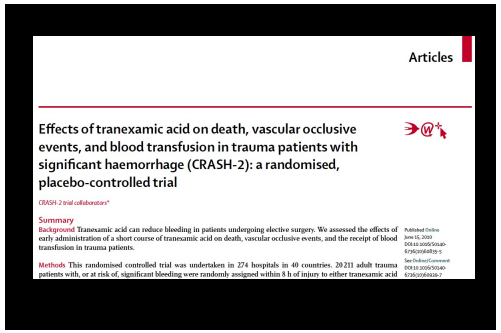
If p is really small (by convention < 5%) you "reject" the null hypothesis, ie, the observed difference is too unusual to just reflect chance variation

The result is "statistically significant"

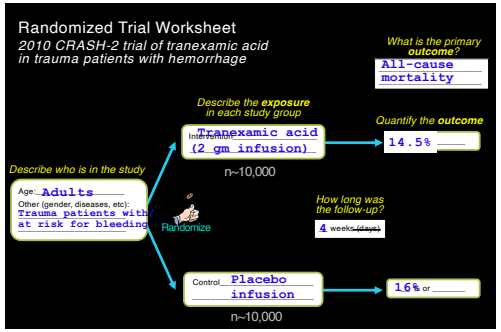
6



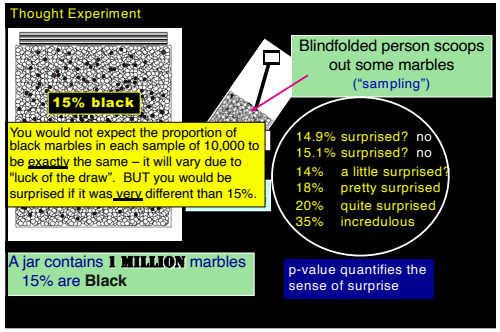
7



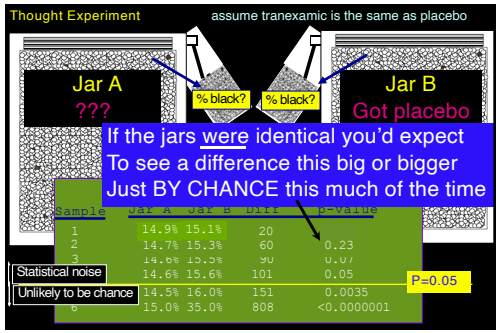
8



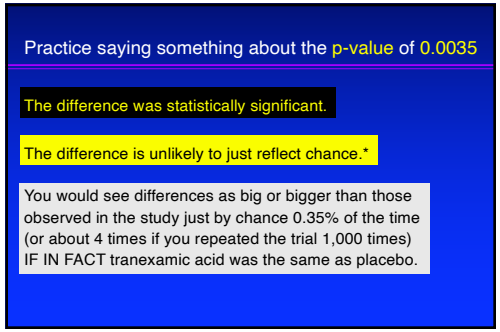
9



10



11



12

What makes p-values higher or lower?

p values are lower with: bigger differences

Study	% Heart attack Placebo	% Heart attack Drug	Absolute difference	Number in each group	p value
A	50%	45%	5%	100	0.48
B	50%	30%	20%	100	0.004

13

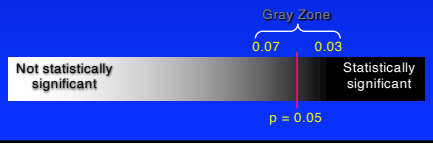
p values are lower with: bigger groups (more outcomes)

Study	% Heart attack Placebo	% Heart attack Drug	Absolute difference	Number in each group	p value
A	50%	40%	10%	20	0.53
B	50%	40%	10%	2,000	<0.0001

14

Now you try:

Women without breast cancer on a low fat diet for 5 years developed fewer cases of breast cancer: 20 vs 22 cases per 1000, p = 0.07



15

p values - Take home points

1. Differences among two groups are the norm.
2. The p value is the probability an observed difference is due to chance.
3. If the p value is > 0.05 – that is 5% - or greater, we say the observed difference may just reflect chance ("not statistically significant").
4. If the the p value is less than 0.05, we say it is unlikely to just reflect chance ("statistically significant").

Advanced point

5. The 0.05 (5%) cut-off is an arbitrary convention (no theories or data to support it). So p values near 0.05 should be interpreted cautiously.

16

p-values are good, confidence intervals are better!

Statistically Significant?

p values (arbitrarily) dichotomize the world (statistically significant or not)

Fewer deaths with tranexamic acid than placebo:
14.5% vs. 16%, p value for the 1.5% difference: p=0.0035

Confidence Intervals (CI's) are a step up because they *also* show you the plausible range of values .

Fewer deaths with tranexamic acid than placebo:
14.5% vs. 16%, CI for the -1.5% difference: -2.5% to -0.5%

17

U.S. Presidential Poll

If the presidential election were held today, who would you vote for?

Biden		Trump		What is B's lead?
50%	-	45%	=	5%

Margin of error version

Margin of error ± 3%

Best guess (lead): 5%

Plausible range: 2% to 8%

Confidence interval version

95% CI: 2% - 8%

Best guess (lead): 5%

Plausible range: 2% to 8%

Confidence interval = Margin of error

Only better – it does the addition/subtraction for you!

18

Randomized trial

How many people were cured?

Bidenex™		Placebo		Absolute difference
50%	-	45%	=	5%

Confidence interval version

95% CI: 2% - 8%

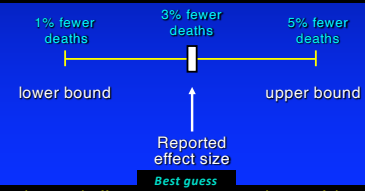
Best guess (difference): 5%

Plausible range: 2% to 8%

19

95% CIs - a visual representation

We are confident that the truth lies within this range (acknowledge that we will be wrong 5% of the time)



The observed effect size is just an estimate of the truth.

20

95% Confidence Intervals

Formal Definition

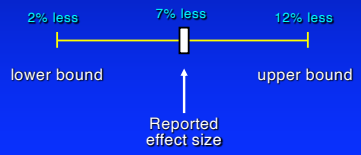
A range of values computed from the sample data which, were the study repeated multiple times, would contain the unknown parameter 95% of the time.

Working definition

95% of the time this range includes the truth.

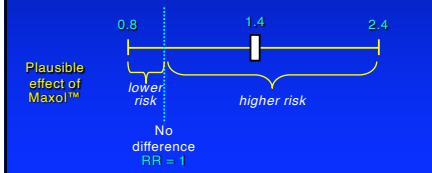
21

Our best estimate is that the intervention reduces cataract surgery **by 7 percentage points**. But this study says it's possible that the intervention may lower surgery by **as little as 2 percentage points** or **as much as 12 percentage points**.



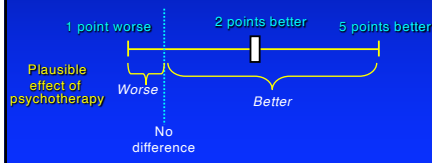
22

The study says it's possible that the risk of polyps may be 0.8 times **lower** with Maxol™ compared to placebo – but it might be 2.4 times **higher.** So, based on this study, we cannot conclude that Maxol™ is different than placebo.

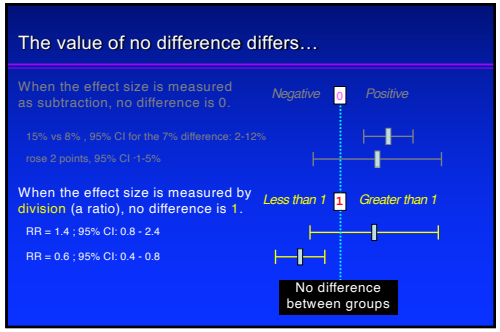


23

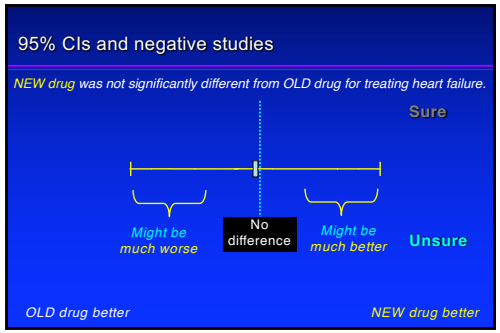
The study says it's possible that depression scores are **worse** by 1 point with psychotherapy – but scores might be 5 points **better.** So based on this study, we cannot conclude that psychotherapy is different than placebo.



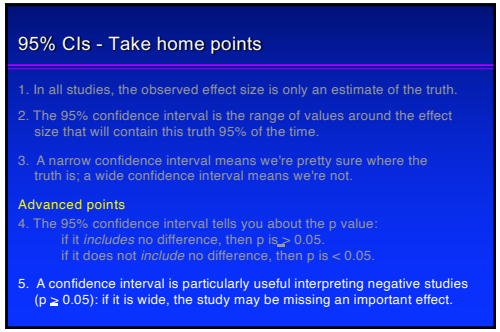
24



25



26



27

Putting it all together:

Children using belt-positioning booster seats had a lower risk of injury than those using seat belts alone:

0.77% vs 1.95%, RR = 0.40 , 95% CI: 0.20 - 0.86

Statistically significant?

Could be due to chance?

Statistically significant, unlikely due to chance, but the apparent effect of booster seats may not be real (may be confounded).

28

Final Caution

A study with a extremely low p value (*very* statistically significant) and an extremely narrow 95% confidence interval (*very* precise) can nonetheless be *very wrong*.

In observational studies, the observed relationship may be *confounded* by other variables.

In all studies, the focus may be on the *wrong patients*, the *wrong comparison* or the *wrong outcome*.

29