

# The Limited Role of Statistics

Steven Woloshin

**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 | The Lisa Schwartz Foundation

1

---

---

---

---

---

---

---

---

### Why findings' might be wrong?

**May not be valid**  
Biased study design or execution

**Just reflect chance**  
Flip a coin 10 times and get 4 heads  
Flip it another 10 times and get 7 heads

Differences are expected

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

2

---

---

---

---

---

---

---

---

### Thought experiment

Exposure

Placebo

Count Death

Placebo

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

---

---

---

---

---

---

---

---

The Limited Role of Statistics  
Steven Woloshin

**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

---

---

---

---

---

---

---

---

# The Limited Role of Statistics

Steven Woloshin

*"Statistically significant"*

p value	Jargon	Interpretation*
$p < 0.05$	The result is "statistically significant"	Really unusual to see such extreme results if null hypothesis true <i>"unlikely to just reflect chance"</i>
$p \geq 0.05$	There is "no significant difference"	Not so unusual to see results this extreme just by chance if null hypothesis is true <i>"may just reflect chance"</i>

**REMEMBER:** p value is just about play of chance — it does NOT tell you if result is "real", true, important

---

---

---

---

---

---

---

---

7

Articles

Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial

*CRASH-2 trial collaborators\**

**Summary**  
Background: Tranexamic acid can reduce bleeding in patients undergoing elective surgery. We assessed the effects of early administration of a short course of tranexamic acid on death, vascular occlusive events, and the receipt of blood transfusion in trauma patients.  
Methods: This randomised controlled trial was undertaken in 274 hospitals in 48 countries. 20,211 adult trauma patients with, or at risk of, significant bleeding were randomised within 3 h of injury to either tranexamic acid.

Received Online: 06 Jun 2010  
DOI: 10.1016/S0140-6736(10)60845-5  
© 2010 Elsevier Ltd  
0950-2688/10/\$ - see front matter  
© 2010 BMJ Group

---

---

---

---

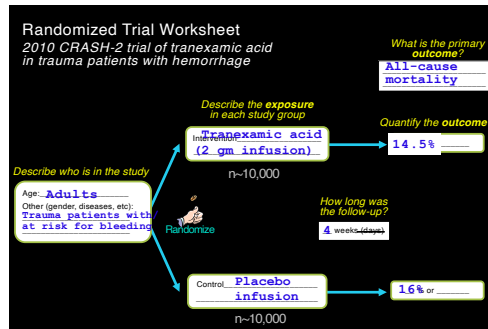
---

---

---

---

8




---

---

---

---

---

---

---

---

9

# The Limited Role of Statistics

Steven Woloshin

Thought Experiment

**15% black**

You would not expect the proportion of black marbles in each sample of 10,000 to be **exactly** the same – it will vary due to "luck of the draw", BUT you would be surprised if it was **very** different than 15%.

Blindfolded person scoops out some marbles ("sampling")

14.9% surprised? no  
 15.1% surprised? no  
 14% a little surprised?  
 18% pretty surprised?  
 20% quite surprised  
 35% incredulous

A jar contains **1 MILLION** marbles  
 15% are Black

p-value quantifies the sense of surprise

---

---

---

---

---

---

---

---

---

---

10

Thought Experiment

assume tranexamic is the same as placebo

Jar A ??? % black? % black? Jar B Got placebo

If the jars were identical you'd expect To see a difference this big or bigger Just BY CHANCE this much of the time

Sample	% black?	% black?	n	p-value
1	14.9%	15.1%	20	
2	14.7%	15.3%	60	0.23
3	14.6%	15.5%	90	0.17
Statistical noise	14.6%	15.6%	101	0.05
Unlikely to be chance	14.5%	16.0%	151	0.0035
	15.0%	35.0%	808	<0.0000001

P=0.05

---

---

---

---

---

---

---

---

---

---

11

Practice saying something about the p-value of 0.0035

The difference was statistically significant.

The difference is unlikely to just reflect chance.\*

You would see differences as big or bigger than those observed in the study just by chance 0.35% of the time (or about 4 times if you repeated the trial 1,000 times)  
 IF IN FACT tranexamic acid was the same as placebo.

---

---

---

---

---

---

---

---

---

---

12

# The Limited Role of Statistics

Steven Woloshin

What makes p-values higher or lower?  
p values are lower with: **bigger differences**

Study	% Heart attack		Absolute difference	Number in each group	p value
	Placebo	Drug			
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		Absolute difference	Number in each group	p value
	Placebo	Drug			
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,  $p = 0.07$

$0.07$        $0.03$   
 Gray Zone  
 Not statistically significant      Statistically significant  
 $p = 0.05$

---

---

---

---

---

---

---

---

15

# The Limited Role of Statistics

## Steven Woloshin

**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 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!**

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%

Statistically Significant?

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  $\pm$  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

---

---

---

---

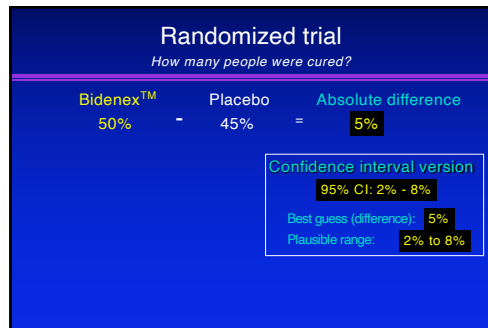
---

---

---

---

The Limited Role of Statistics  
Steven Woloshin



---

---

---

---

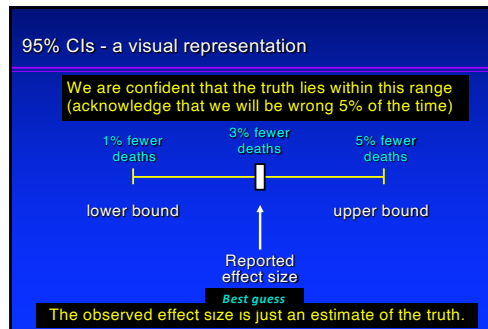
---

---

---

---

19



---

---

---

---

---

---

---

---

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.

---

---

---

---

---

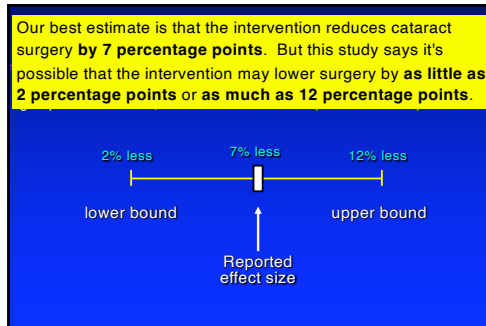
---

---

---

21

The Limited Role of Statistics  
Steven Woloshin



22

---

---

---

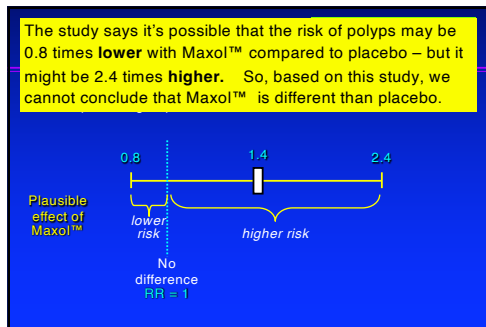
---

---

---

---

---



23

---

---

---

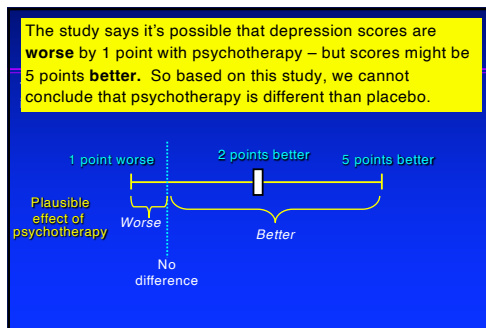
---

---

---

---

---



24

---

---

---

---

---

---

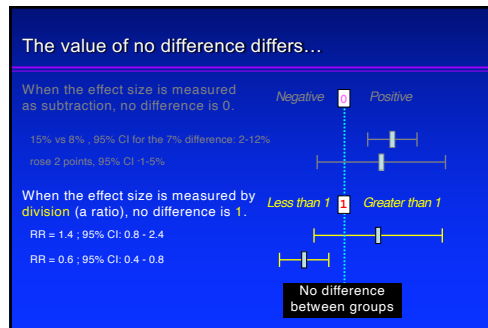
---

---



# The Limited Role of Statistics

Steven Woloshin



25

---

---

---

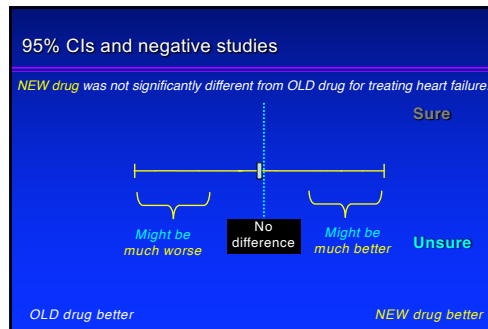
---

---

---

---

---



26

---

---

---

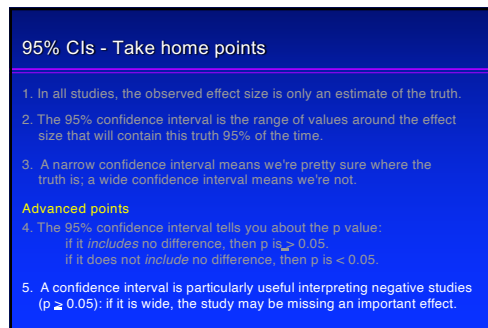
---

---

---

---

---



27

---

---

---

---

---

---

---

---

The Limited Role of Statistics  
Steven Woloshin

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