Statistically significant versus patient-importance

Aleksi Reito, MD, PhD

O&T resident, assoc prof

Tampere University Hospital

Disclosure

- Salary
  - Pirkanmaa Hospital District, Pihlajalinna
- Writing fee
  - Finnish Medical Journal
  - Duodecim
- Consultation fee
  - University of Jyväskylä

Content

- Recap of p-values and confidence intervals
- Use of p-value and confidence intervals
- Interpretation of study findings
- Implications of study findings

From the introductory statistics…

- *"Statistically significant is not the same as clinically significant"*
  - In large samples small differences or association become statistically significant
- Statistically non-significant can be clinically significant

Statistically significant?

- Frequentist statistics
  - Repeated sampling from a population
  - P-values, confidence intervals, power (type II error), type I error, alfa level, effect size
- Basis of our science = p-value < 0.05
  - Statistically significant! Success!!

P-value?

- Probability of obtaining as extreme or more extreme result in a repeated identical study
- Mean difference in VAS pain 12/100, p=0.12
  - ≠Probability of true finding
  - ≠Probability of chance

How do we use p-values?

- **P-value < 0.05**
  - **Statistically significant findings → huge success!**
  - ***Treatment has an effect***
  - ***Variables had an association***
- ***Statistically significant is not the same as clinically significant***
  - ***Mean difference in blood loss was 40 mL ($p<0.05$)***
  - ***Mean difference in VAS pain was 9/100 ($p<0.05$)***

How do we use p-values?

- **P-value > 0.05**
  - **Statistically insignificant finding → too bad, no publication for you**
  - ***Treatment does not have an effect…***
  - ***Treatments are equal…***
  - ***Variables do not have an association…***
  - ***We found no evidence…***

P-values only provide evidence against null hypothesis

- Hypothesis testing
  - Null hypothesis
    - $\mu_1 = \mu_2$, RR = 1
  - Alternative hypothesis
    - $\mu_1 \neq \mu_2$, RR ≠ 1
- <u>P-value >0.05 does not mean that the null hypothesis is true</u>
  - We can never state that groups were equal or non-different
  - ***Absence of evidence is not evidence of absence***
- False conclusion with major ethical and safety aspects

No effect statements are paradoxical

- Surgical procedure → there was no effect (?)
- Vasodilating drug → there was no effect (?)
- Salt water for pancreatic cancer → probably no effect
- <u>There is always an effect, it is only a matter of sample size whether we can detect it</u>
  - Insulin for type 1 diabetic
  - Joint lavage in osteoarthrosis

P-value based reasoning is dichotomous

- Risk ratio 0.84 (not significant or $p>0.05$)
  - ”We observed no effect…”
- Risk ratio 0.84 (significant or $p<0.05$)
  - ”We observed an effect…”
- Yes or no answers are unintuitive
  - Nothing magical happens at $p=0.05$

ANDROMEDA

JAMA Network™

QUESTION Does a resuscitation strategy targeting normalization of capillary refill time, compared with targeting serum lactate levels, reduce mortality in patients with septic shock?

CONCLUSION This randomized clinical trial of adults with septic shock found that use of a peripheral perfusion-targeted resuscitation strategy, compared with targeting serum lactate, did not significantly reduce mortality.

POPULATION

198 Men 226 Women

Adults in the ICU with septic shock

Mean age: 63 years

LOCATIONS

28 ICUs in 5 countries in South America

INTERVENTION

424 Patients randomized

212 **Peripheral perfusion group**
Resuscitation protocol of normalizing capillary refill time (measured in seconds)

212 **Lactate group**
Resuscitation protocol of normalizing or decreasing lactate levels (>20% per 2 hours)

PRIMARY OUTCOME

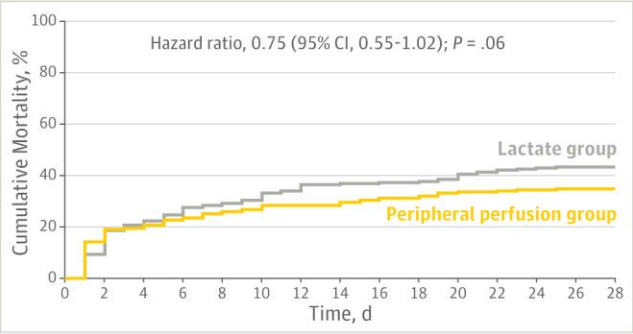
All-cause mortality at 28 days

FINDINGS

All-cause mortality at 28 days

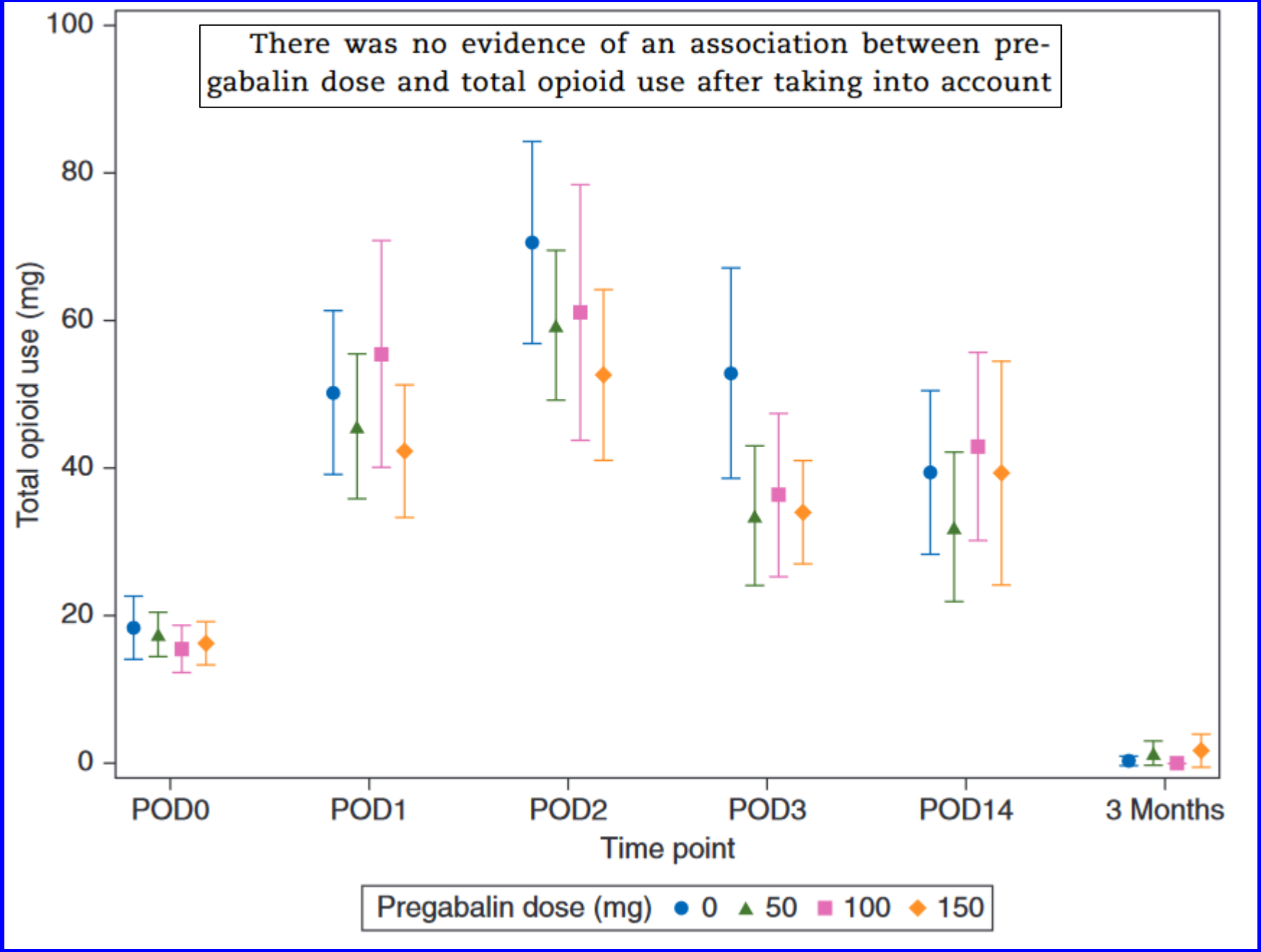
Peripheral perfusion group 34.9% (74 patients died)

Lactate group 43.4% (92 patients died)



No significant risk difference between groups: -8.5% (95% CI, -18.2% to 1.2%),

Hernández G, Ospina-Tascón GA, Petri Damiani L, et al. Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock: the ANDROMEDA-SHOCK randomized clinical trial [published February 17, 2019]. *JAMA*. doi:10.1001/jama.2019.0071


There was no evidence of an association between pregabalin dose and total opioid use after taking into account
Total opioid use (mg)
100
80
60
40
20
0
POD0
POD1
POD2
POD3
POD14
3 Months
Time point
Pregabalin dose (mg) 0 50 100 150

What then?

- As said there is always on effect
  - Is it precise enough to declare it is (statistically/clinically) significant

Confidence interval to the rescue

- Confidence intervals should always be used in the interpretation of study results
  - CIs convey information about (im)precision!
  - RR 0.84 (95% CI: 0.35 – 2.01)
  - RR 0.84 (95% CI: 0.66 – 1.08)
  - RR 0.84 (95% CI: 0.73 – 0.97)

Confidence interval?

- Range of values which will include true population value 95% of the time
  - Same study or test performed 100 times
- RR 0.84 (95% CI: 0.73 – 0.97)
  - Probability of including 0.84 in the next studys CIs ~83.4% (1)
- CI is not a probability → go Bayesian

Confidence interval?

- Uncertainty intervals (2)
- Compatibility intervals (3)

HEAD TO HEAD

Are confidence intervals better termed "uncertainty intervals"?

Debate abounds about how to describe weaknesses in statistics. **Andrew Gelman** has no confidence in the term "confidence interval," but **Sander Greenland** doesn't find "uncertainty interval" any better and argues instead for "compatibility interval"

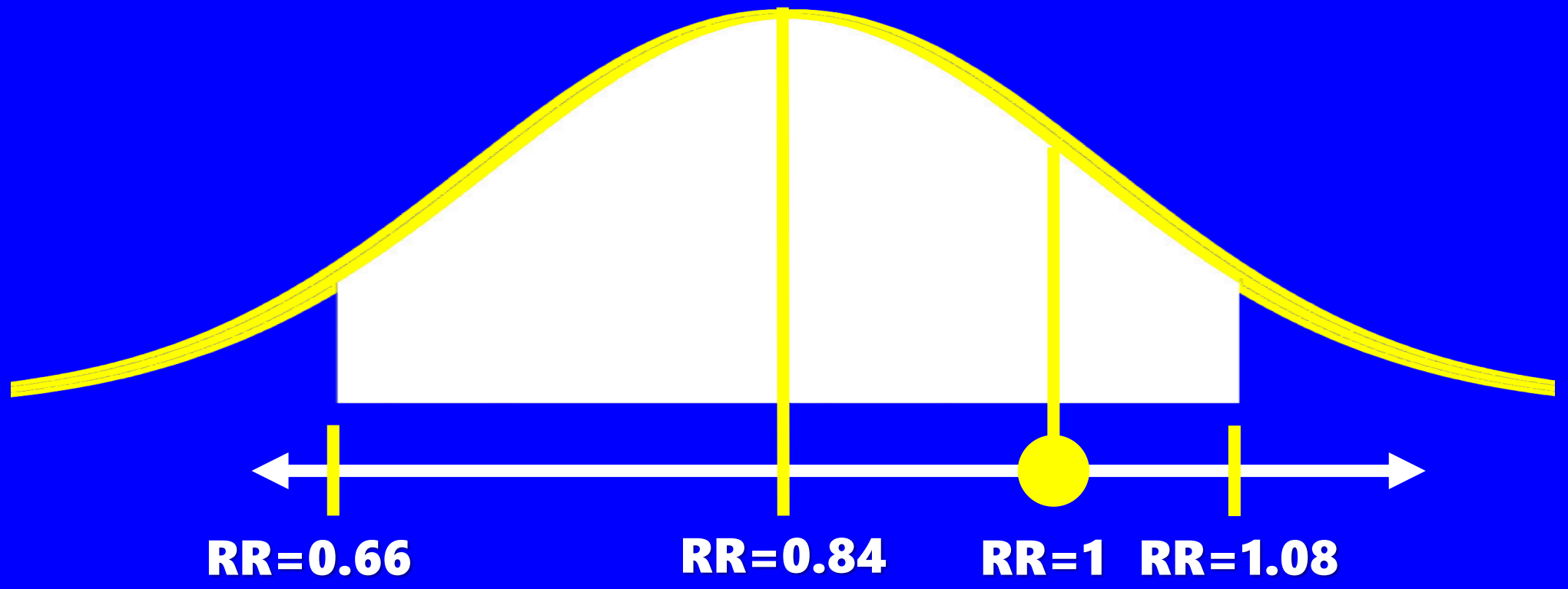
Andrew Gelman *professor of statistics and political science*[1], Sander Greenland *professor of epidemiology and statistics*[2]

Compatibility interval

- Risk ratio 0.84 (95% CI: 0.66 – 1.08)
  - Data available is <u>compatible</u> with 34% reduction and 8% increase in the risk of death/complication/… etc

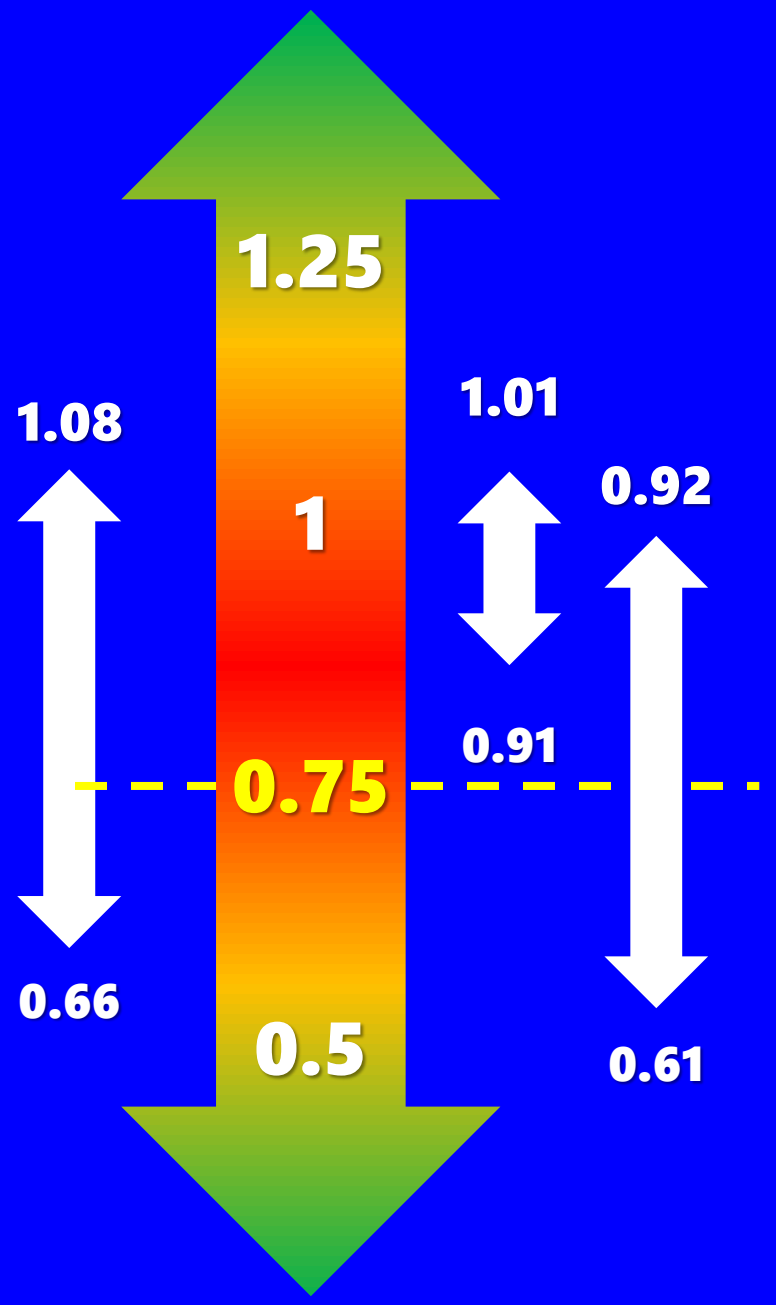
Compatibility interval

- Risk ratio 0.84 (95% CI: 0.66 – 1.08)



What can be said?

- Risk ratio 0.86 (95% CI: 0.66-1.08)
- Risk ratio 0.97 (95% CI: 0.91-1.01)
- Risk ratio 0.75 (95% CI: 0.61-0.92)
- What are the relevant values in terms of the confidence intervals?



What is relevant and for whom?

- Minimal clinically important difference (MCID)
  - According to who?
- Patient important
  - 250 mL blood loss?
  - 15/100 reduction in VAS pain?
- Population important
  - Risk reduction of 10%
- Humanly important (4)

With great power comes great responsibility

- Any study conclusion comes with a value-set
  - Unless: *(No) effect was seen…*
- Well outlined study conclusion considers the result uncertainty
- Uncertainty must be judged against some values
  - What is important/relevant/human?

With great power comes great responsibility

- Risk ratio = A (95% X-Y)
  - <u>What does the X and Y include and exclude?</u>

From a population to an individual

- Patient important not so easy
- Average treatment effect
  - In a population alike included in the trial
  - ”On average patients (do not) improve….”
- Point estimate and CIs cannot be applied to an individual patient

From a population to an individual

- We can never estimate the outcome in a certain individual using a single study result
  - ☹ Will this patient have patient-important improvement?
  - ☺ If I treat this population, will I achieve benefit?
- Outcomes in an individual patient can be estimated conditionally using available baseline variables (5)
  - Risk models, clinical prediction models
  - Is the estimated improvement patient-important?

To conclude

- Study results are never dichotomous
- Statistical significance does not really matter
- Each study results has some level of uncertainty and imprecision
- All robust study conclusions are based on some value of patient/population/humanly importance

References

1. Cumming G, Maillardet R. Confidence intervals and replication: where will the next mean fall? Psychol Methods. 2006 Sep;11(3):217-27.
2. Gelman A, Greenland S. Are confidence intervals better termed "uncertainty intervals"? BMJ. 2019 Sep 10;366:l5381
3. Rafi Z, Greenland S. Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. BMC Med Res Methodol. 2020 Sep 30;20(1):244.
4. Guyatt G et al. Patients at the center: in our practice, and in our use of language. ACP J Club. Jan-Feb 2004;140(1):A11-2.
5. Kent DM et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern Med. 2020 Jan 7;172(1):35-45.

Thanks! Questions?

- **Email: aleksi@reito.fi**
- **Slides available → reito.fi/luennot**
- **Twitter: *@AleksiReito***