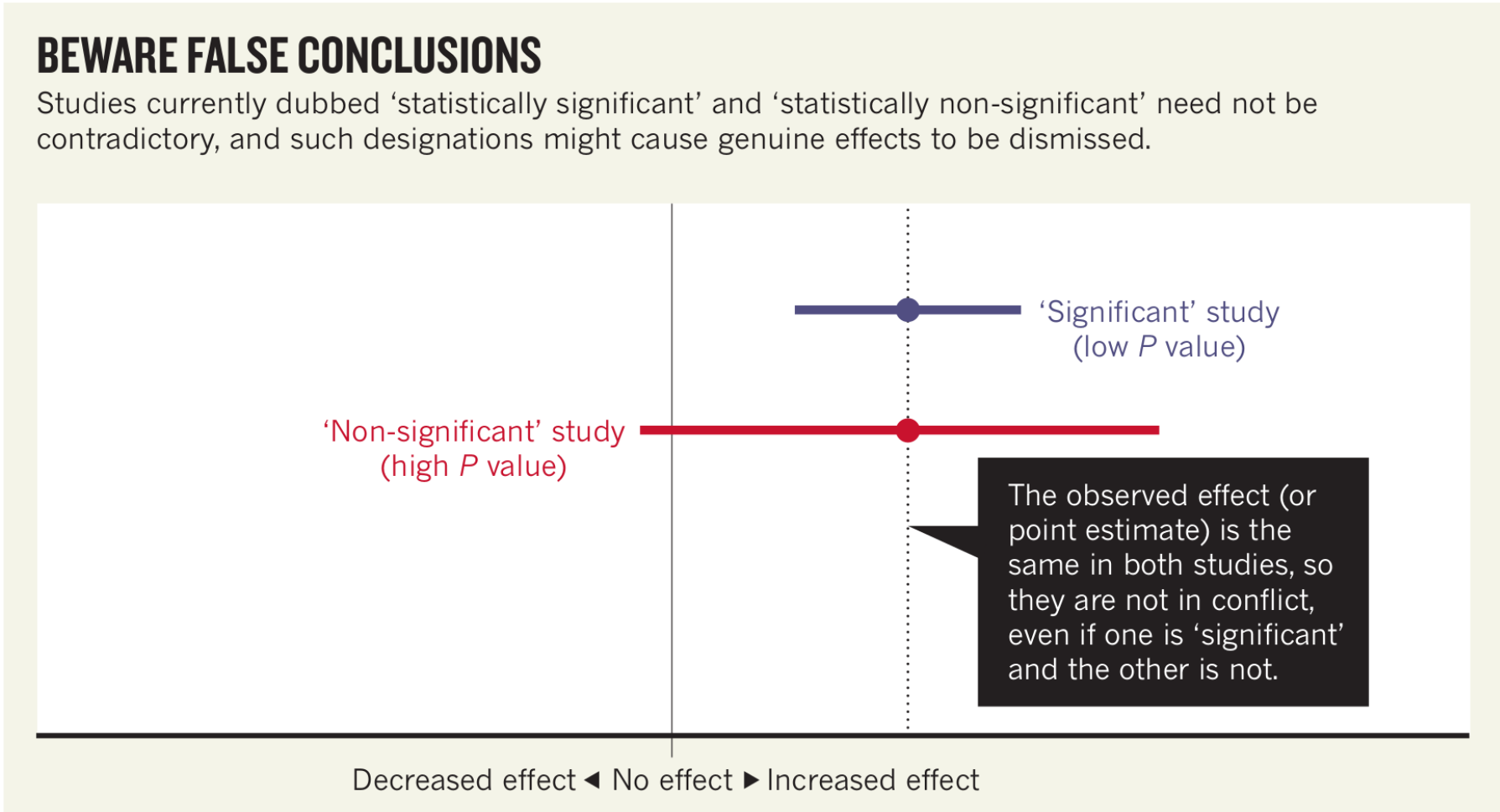
I will draw some 95% confidence intervals

. . .

Look at each of the figures

- Which one(s) will be significant (at 0.05)?
- Which are compatible with strong positive effects?
- Which ones make it sensible to try to focus on this gene as a possible target?
  - what if chasing this lead is extremely expensive or risky (side effects?)?
  - what if it is cheap?
- Which one(s) make it sensible to forget about this gene as a possible target?
- We are using 95%. What if we used 99%? Or 90%?



From Amrhein et al., 2019. *Retire statistical significance*, Nature, 305.

Hypothesis vs. significance testing

A. How can these data be interpreted?

APPROACH	QUESTION	STATISTICAL ANALYSIS	INTERPRETATION
Hypothesis test ("bright line")	Should we act as though the observed effect is nonzero (given prespecified error rates)?	1. $P \leq 0.05$ 2. $P \leq 0.05$ 3. NS	Studies 1 and 2 indicate action based on a nonzero true effect is justified. Study 3 indicates it is not.
Fisherian *P* value	How much evidence is there that the true effect is different from zero?	1. $P = 0.03$ 2. $P = 0.05$ 3. $P = 0.11$	Studies 1 and 2 provide moderate, statistically significant evidence that the new treatment is better. Study 3 supplies weak but insufficient evidence to say the treatment is effective.
Estimation	What range of true effects is statistically consistent with the observed effects?	Effect, 95% confidence interval (%) 1. 6, 0.5 to 12 2. 20, 2.5 to 38 3. 6, –1.4 to 13	Studies 1 and 3 indicate the new treatment had a small to moderate effect. Study 2 is consistent with either small or large effects.

From Goodman, 2016, *Aligning statistical and scientific reasoning*, Science, 352.

A particular model, some assumptions

- We are making assumptions
- There is an underlying model
- P-value, CI, constructed based upon the model and its assumptions.

- Decision: include costs of mistakes (more later)
- What p-value? Depends on context. Sometimes we might want it much, much, much smaller.

Context is key

> “Dichotomous conclusions can be useful for pinning down discoveries of gene variants for osteoporosis, new bosons or carcinogens, say.
>
> But focusing on effect sizes can often be better than determining whether an effect exists.”

Ioannidis, letter to Nature, 28 March, 2019, vol. 567.
“Retiring significance: a free pass to bias”

Taking someone to trial and presumption of innocence analogy

Significance and hypothesis testing Is there enough evidence to sent this person to jail? How likely is, if she were innocent, that she committed the crime?

Estimate But I really think she is guilty. Or “but the evidence is most compatible with her being guilty (though we cannot show it beyond reasonable doubt)”.

Bioquivalence You have to show your innocence.

Let’s go back to the initial cases

. . .

Look at each of the figures

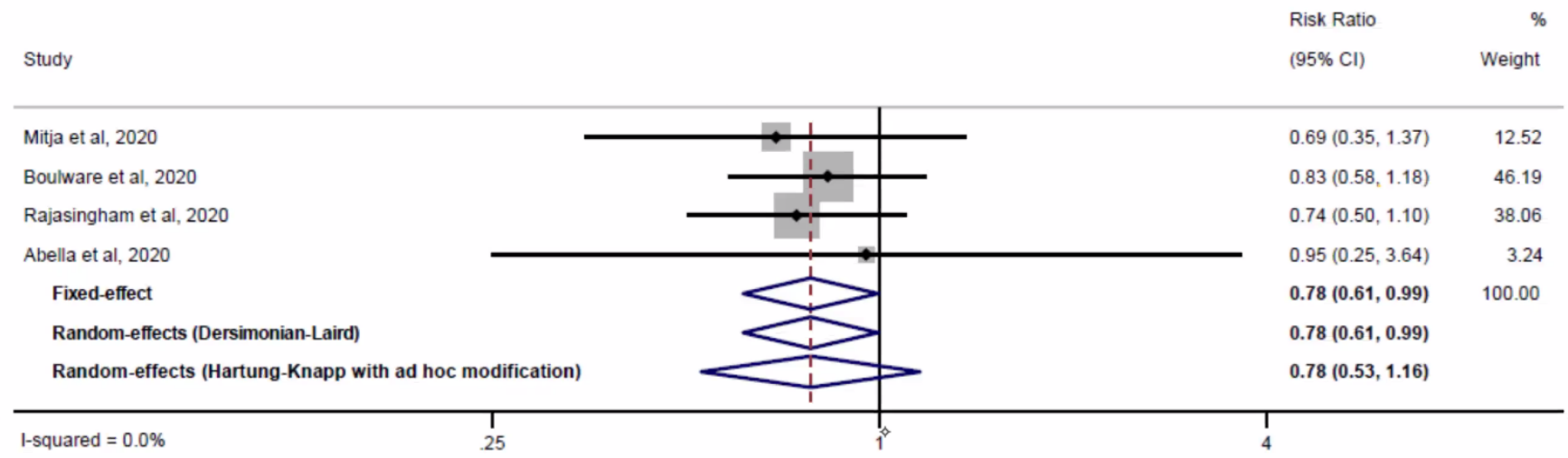
- Which one(s) will be significant?
- Which are compatible with strong positive effects?
- Which ones make it sensible to try to focus on this gene as a possible target?
  - what if chasing this lead is extremely expensive or risky (side effects?)?
  - what if it is cheap?
- Which one(s) make it sensible to forget about this gene as a possible target?
- We are using 95%. What if we used 99%? Or 90%?

Does all of this matter?

- A lot!!
- And what about wide CIs?
- Many studies seriously misinterpreted. It has happened a lot, with serious consequences with the covid pandemic.

The Hydroxychloroquine story
4 randomized trials of HCQ for prophylaxis

- Garcia-Albeniz et al. medRxiv 2020 (November version)



From Miguel Hernán's talk, Causal inference, "Epidemiological Research Methods in the COVID-19 Pandemic", IAP, 2021-06,14. Also García-Albéniz et al., 2021, https://www.medrxiv.org/content/10.1101/2020.09.29.20203869v4.full

Wide CIs: problems and consequences of misinterpretations

- Many papers published with huge CI are papers with no information, basically
- But then, that leads to people believe that the drug is not effective
- That makes further additional studies unlikely or impossible because "we already know it does not work" (which is the wrong inference)

References

- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond p < 0.05. *The American Statistician*, 73(sup1), 1-19. https://doi.org/10.1080/00031305.2019.1583913
- Stats blogs. For example https://statmodeling.stat.columbia.edu/2019/03/20/retire-statistical-significance-the-discussion/
- Goodman, S.N. (2016). Aligning statistical and scientific reasoning. *Science*, 352(6290), 1180-1181. https://doi.org/10.1126/science.aaf5406
- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance ("Retire statistical significance"). *Nature*, 567(7748). https://doi.org/10.1038/d41586-019-00857-9
- Greenland, S., et al. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31(4), 337–350.* https://doi.org/10.1007/s10654-016-0149-3
- García-Albéniz et al. (2021). Systematic review and meta-analysis of randomized trials of hydroxychloroquine for the prevention of COVID-19. *medRxiv*. https://www.medrxiv.org/content/10.1101/2020.09.29.20203869v4.full