Two small-data $n \ll p$ case studies: hydroxychloroquine, Marseilles and Meijel

High dimensional data – One World Symposium

Thursday 27 August, 2020

Richard Gill
Leiden University, Netherlands
https://www.math.leidenuniv.nl/~gill

Didier Raoult

Gautret et al. (2020), Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents* **56** 105949 (6pp.)

https://rpubs.com/gill1109/raoult

Rob Elens, MD

Family doctor and alternative medicine practitioner, Meijel, Netherlands

Dutch TV appearances, Dr Elens is a figurehead of a popular pro-HCQ movement

https://rpubs.com/gill1109/elens

Marseilles

	good outcome	bad outcome
treatment	15	11
control	2	14

`p-value = 0.004491`

Meijel

	good outcome	bad outcome
treatment	13	13
control	0	10

`p-value = 0.005848`

Gautret et al. (2020)
6 patients removed from treatment group
for “non-compliance”

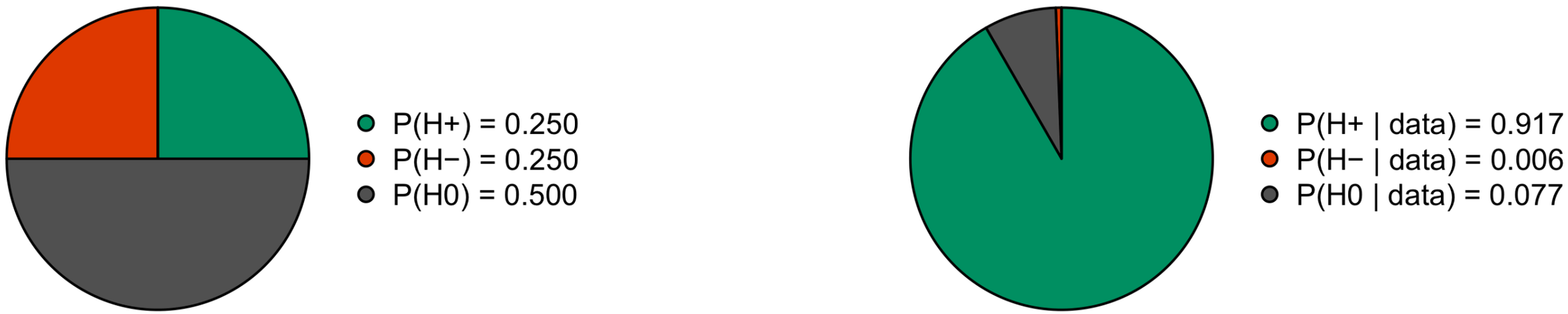
Remaining: 14, 6; 2, 14
p-value = 0.001

Fisher exact tests

“Outcome” = Covid-19 virus absent/present after 6 days

The Bayesian conclusion (Marseilles)

JASP; AB test (R)

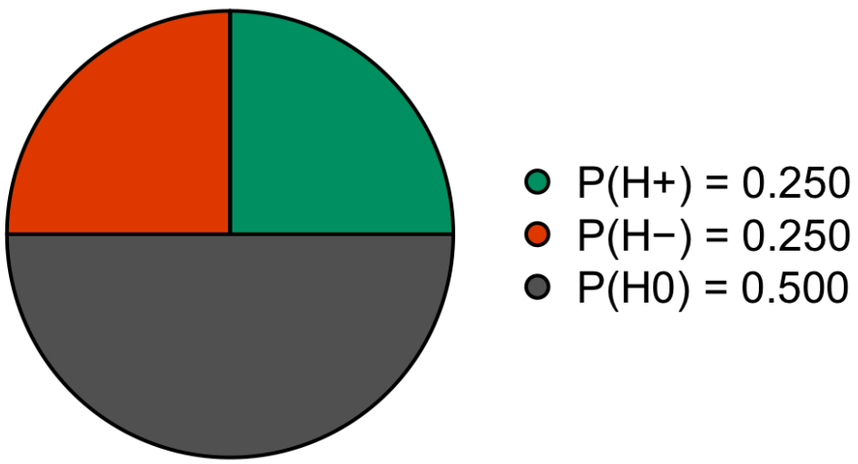


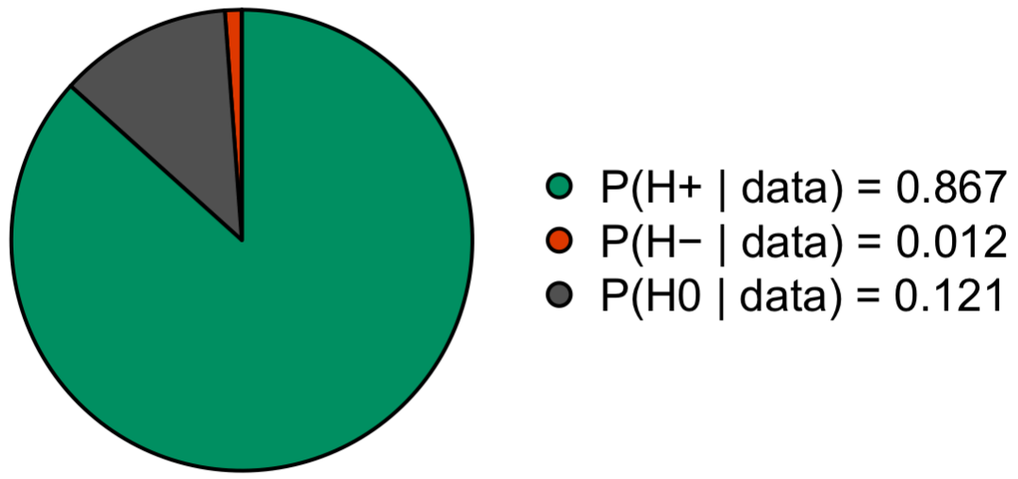
Left: prior; right: posterior

A posteriori, still 8% chance of no difference!

The Bayesian conclusion (Meijel)

JASP; AB test (R)





Left: prior; right: posterior

A posteriori, still 12% chance of no difference at all!

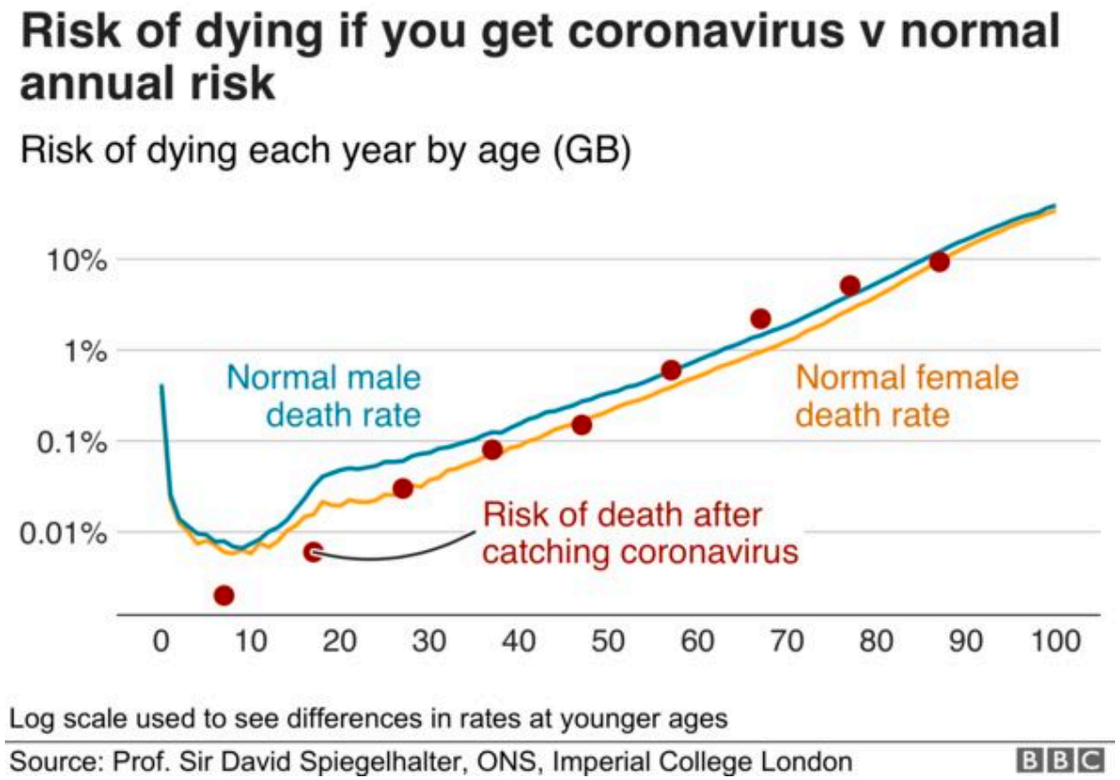
Confounders

Data = two ca. 40 x 40 spreadsheets

- Treatment (binary)
- Outcome (disease free after 6 days)
- Age (from 20 to 90); sex
- Comorbidities (obesity, diabetes, ... : yes/no)
- Symptoms (sense of smell gone, fever, ... : yes/no)
- Also some other numerical variables (blood pressure, ...)

First findings

- "Common sense" and medical knowledge reduces # confounding variables to half a dozen
- Logistic regression on all 6 gives nonsense
- Lasso on all 6 gives nonsense
- Logistic regression with just one or two covariates shows realistic (significant) coefficients; effect of age (ages range from 20 to 90) exactly what we expect … just like a life table



Present experiments

- Compute a “standardised age” from mortality statistics (risk of death this year, given age, sex, comorbidities)
- Compute a “symptoms severity index” – we can now do this using much bigger published data sets!
- Now we have just two continuous covariates (use simple spline curve) and one discrete. Results coming very soon, I hope!

Hoang Van Thuan, Marseilles and Vietnam

Leila Schneps, Paris

Dipro Mondal, Leiden