There is nothing as practical as a good philosophy

Daniel Lakens, Eindhoven University of Technology, Holland

Déjà eu. How we keep on losing and re-inventing statistical theory '.

Stephen Senn, University of Sheffield and the Medical University of Vienna

Comments by Clelia Di Serio

Laken's Talk

Yes, how data should be analyzed depends, but in some cases, it primarily depends on your philosophy of science

- Addendum: how data should be analyzed depends on «statistical reasoning», how a statistician evaluates EVIDENCE.

1. HOW MUCH EVIDENCE THERE IS IN DATA?

- Do we really know our data?

- Statistical Reasoning

What is statistical Reasoning?

- shadows = data,
- cat/table/chair = true model, true parameters
- flickering fire = random noise, sampling, ascertainment bias, confounding

Statistical reasoning is controlling re so prisoners learn real objects only from shadows.

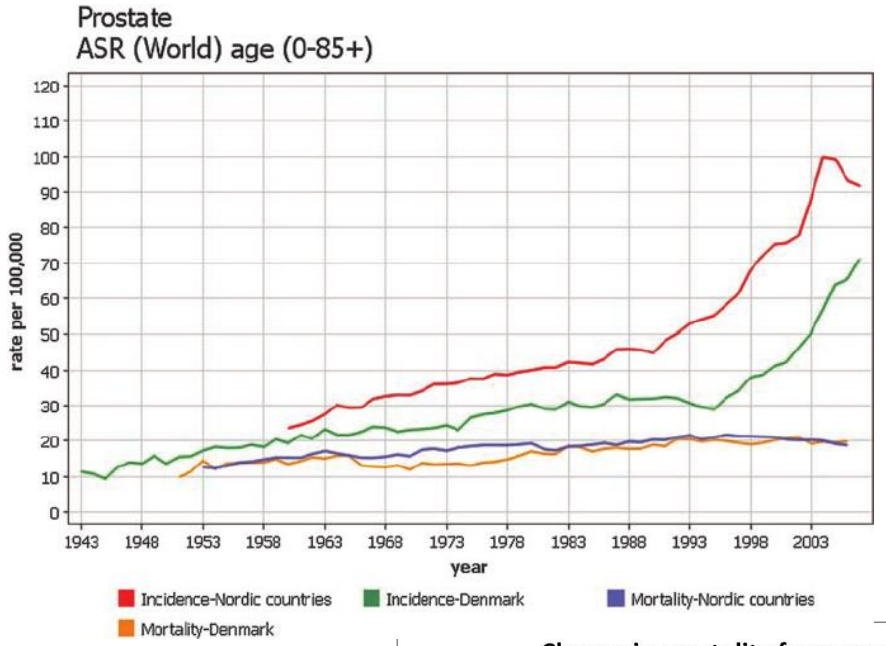
What are data?

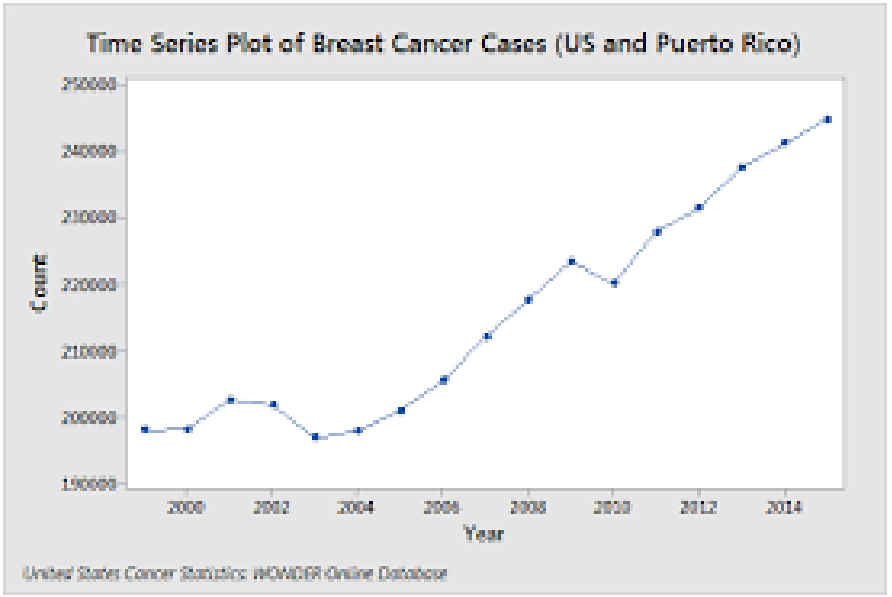
DATA ARE SHADOWS…..IF WE DO NOT UNDERSTAND THE «DATA GENERATING PROCESS WE CANNOT DO INFERENCE.
ISSUE OF REPRESENTATIVENESS!

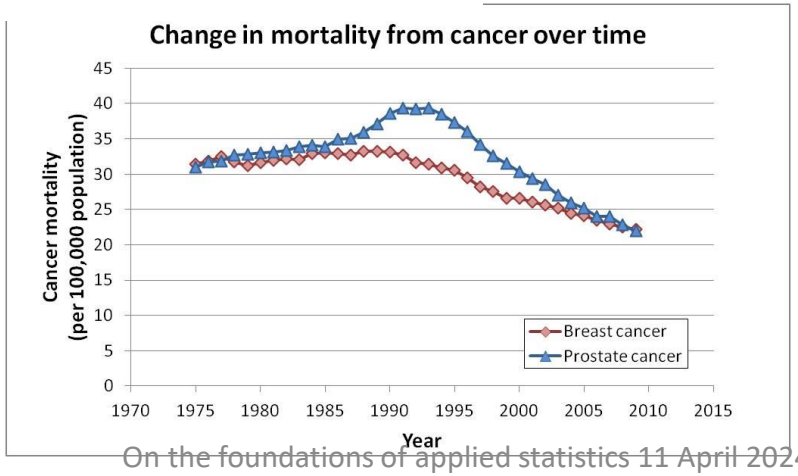
Do we really measure noise in data generating process?

- Are our «measures» replicable «over time»?
- Effect of technology on estimates in longitudinal series….

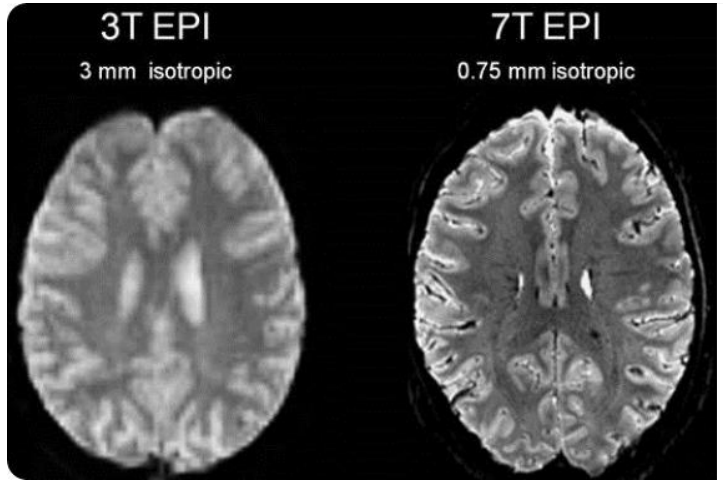
Do we consider/measure noise?

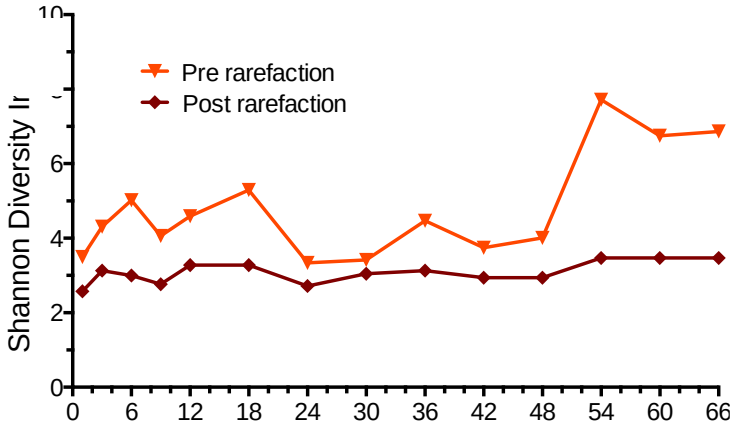






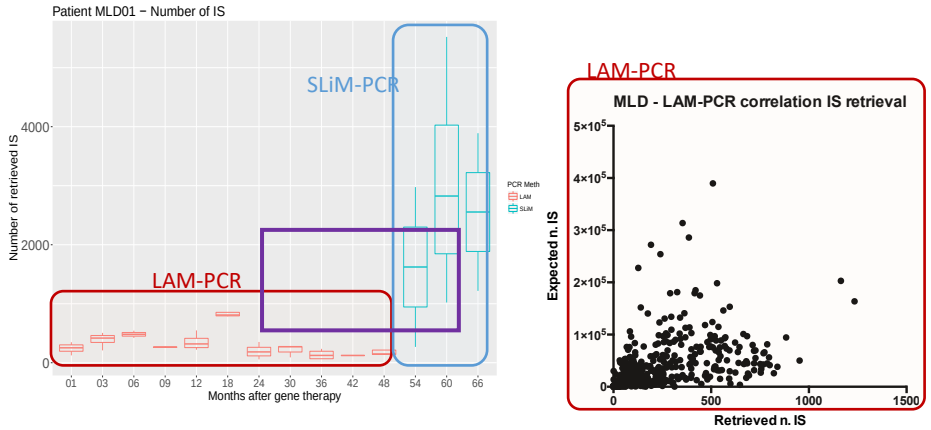
How much noise in technology change?





Gene therapy clinical trial, example from MLD

- Effects of introducing new PCR methods for integration sites (IS) identification: LAM-PCR → SLiM-PCR
  - Higher efficiency in data retrieval (>5-20 fold)
  - Higher accuracy and precision in clonal quantification
  - Requires new methods for data harmonization and integration in aggregated measures



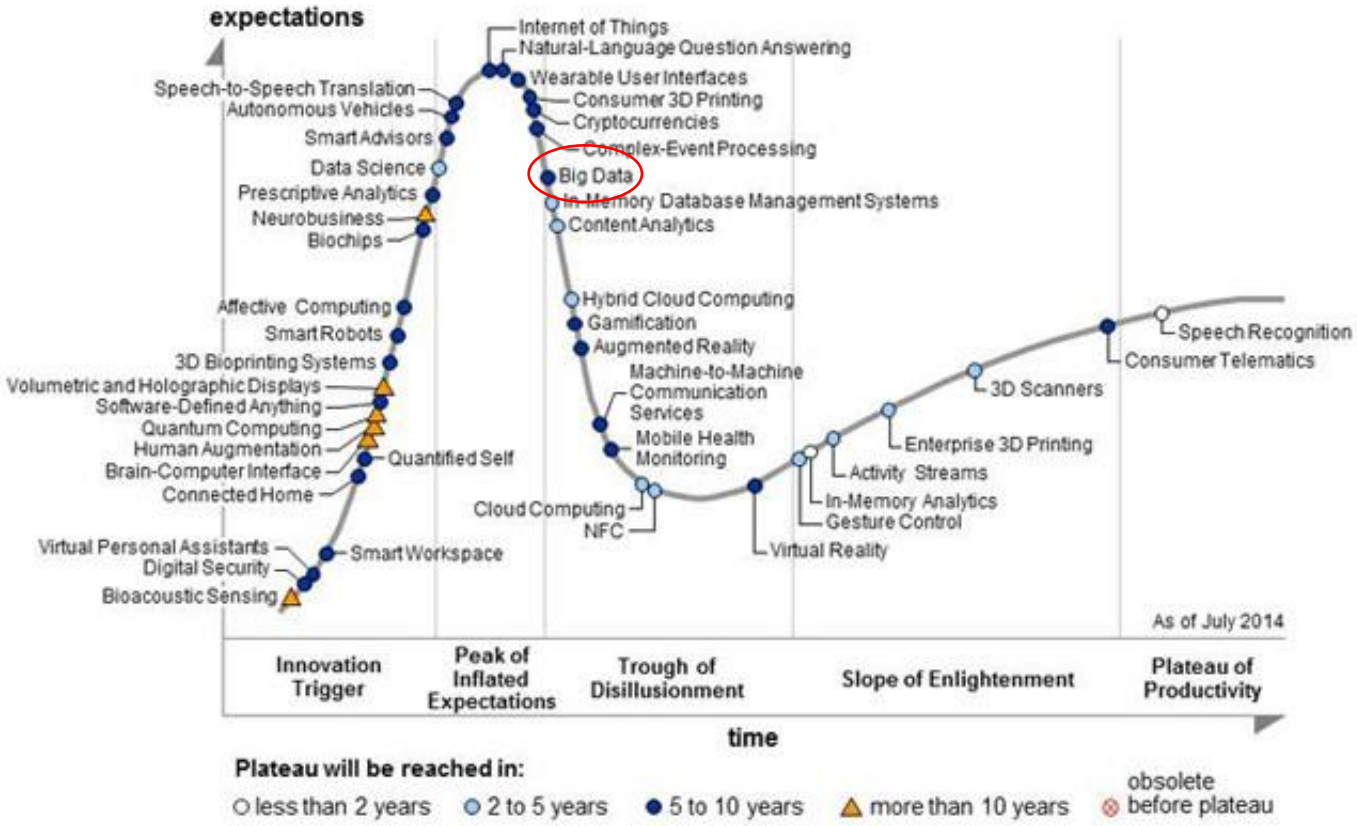
Senn's Talk

- The difference between agricultural and medical research is that agricultural research is not done by farmers

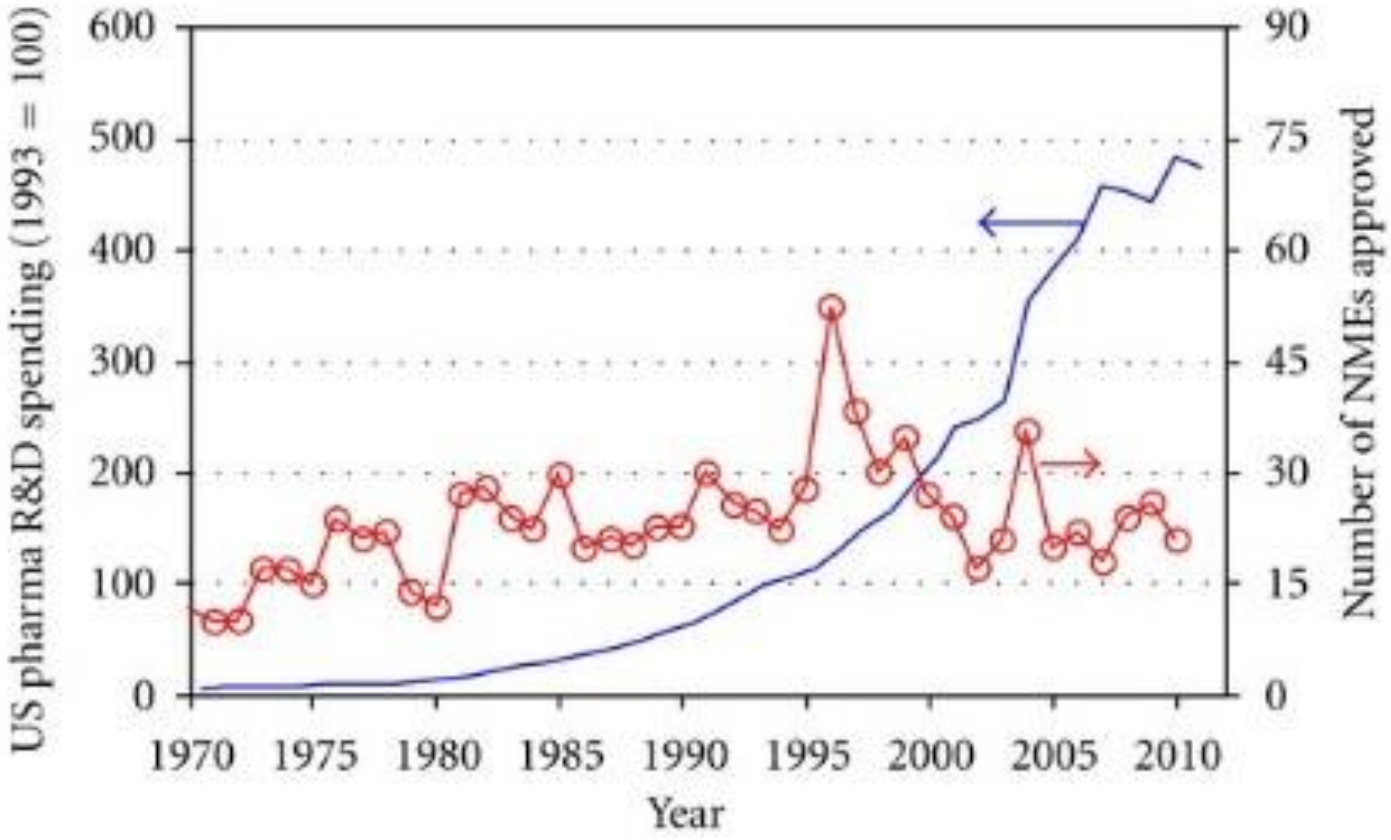
Michael Healy

Reproducibility e replicability crisis

Big data in medicine in 2024



Research Costs (pharma)…VS…Number of New Medical Entities (discoveries)

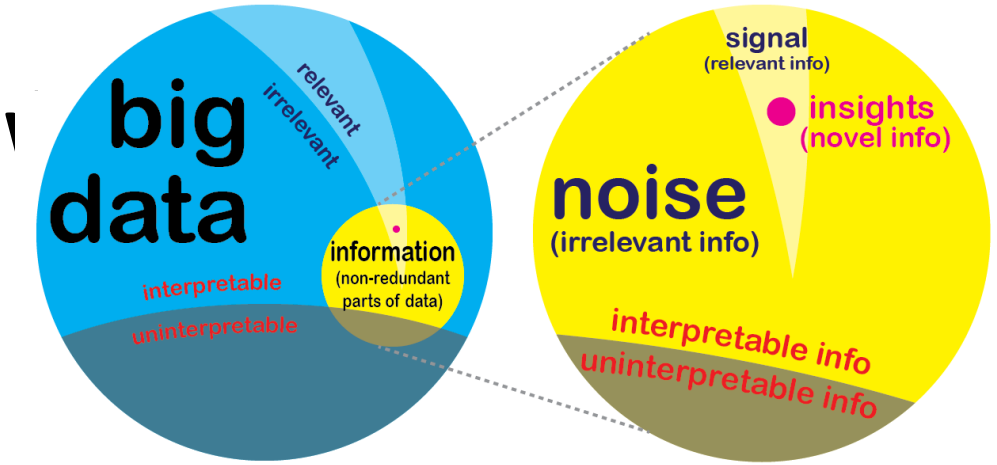


BIG DATA: VVV or WWW

WWW.BIGDATA.COM

W=who ; W = when; W = what

BIG DATA = REPRODUCIBILITY PROBLEMS IN HEALTH.

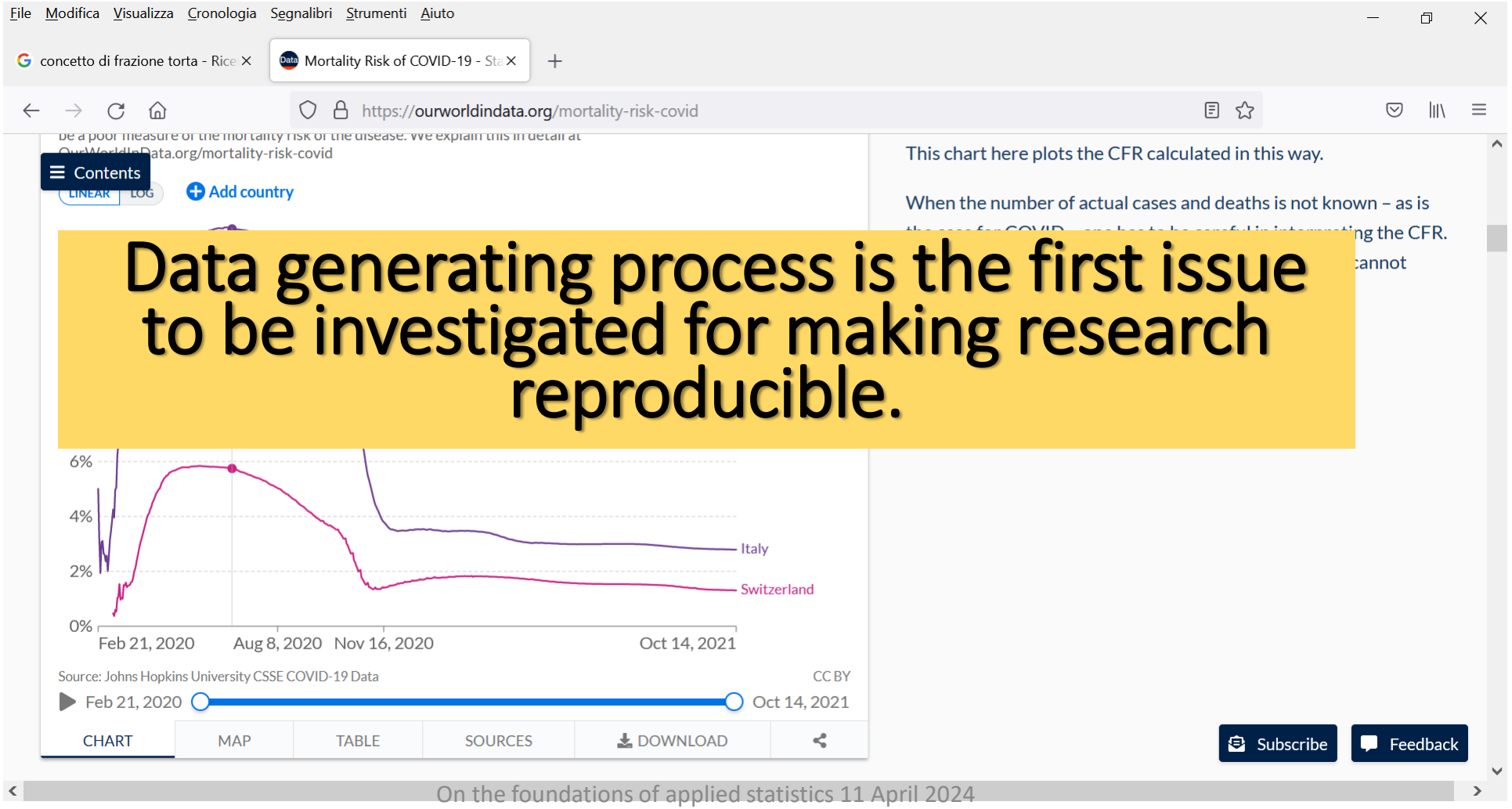


REPRESENTATIVE SAMPLING?

IMPROVE INFERENCE?

BIG DATA= BIG INFORMATION?

HOW MUCH NOISE WHEN WE DO NOT CONTROL DATA GENERATING PROCESS?



Major hurdle in modern INFERENCE: Lack of reproducibility in the results!

Reasons?

LACK OF DESIGN! (RWD)

1. Lack of design: Poor data quality. No check on information quality.
2. Lack of design: Wrong measures to use evidence
3. Lack of design: Artificial Intelligence tools used with unknown data-generating process: lack of representativeness
4. Need of new scheme of design and sampling strategies