Data, models and science

Jonathan Dushoff, McMaster University

1 Introduction

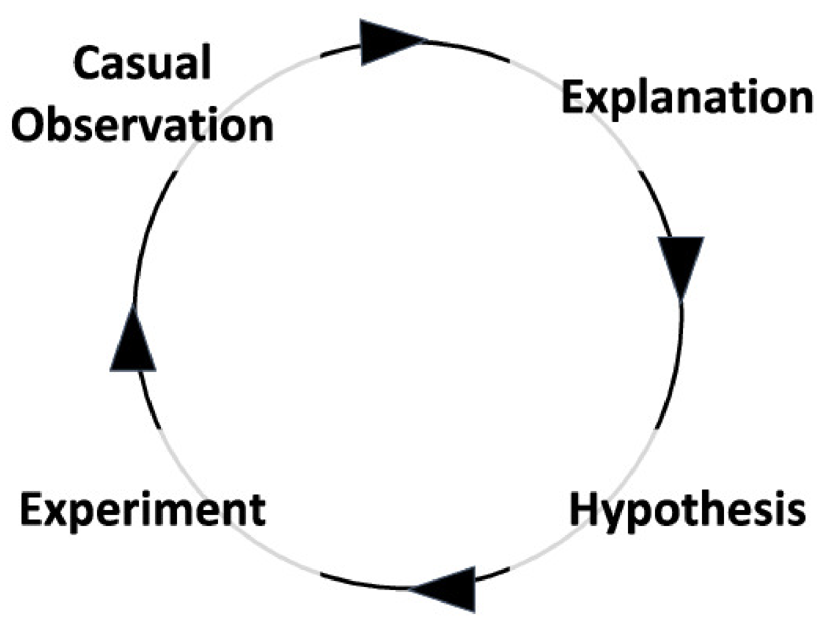
Goals

- Process of science
- How science informs public health
  - Specific examples
- Approaches to epidemiology

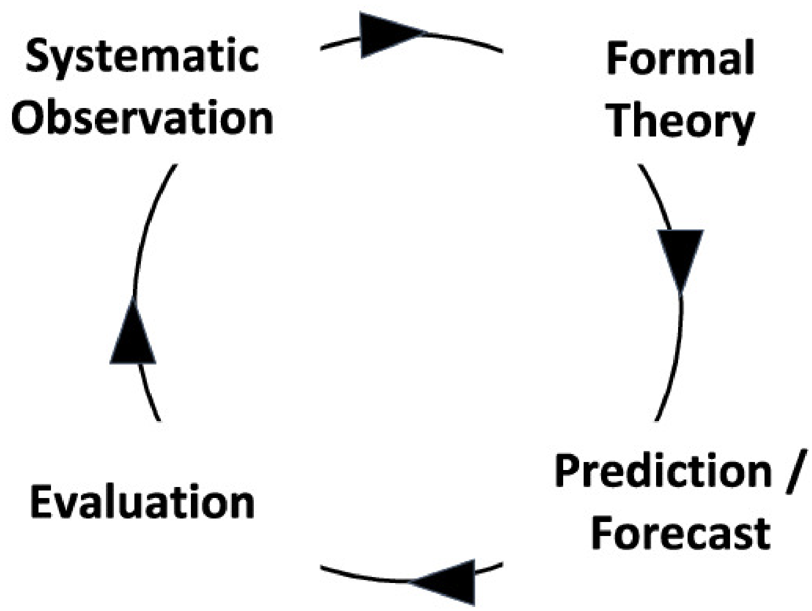
Science is a *process*

- Observe and experiment with reality to *discover* and *challenge* ideas about how it works
- A key to science is that everything is open to question
  - Science is the belief in the ignorance of experts – *Feynman*

The process of science



Science without experiments

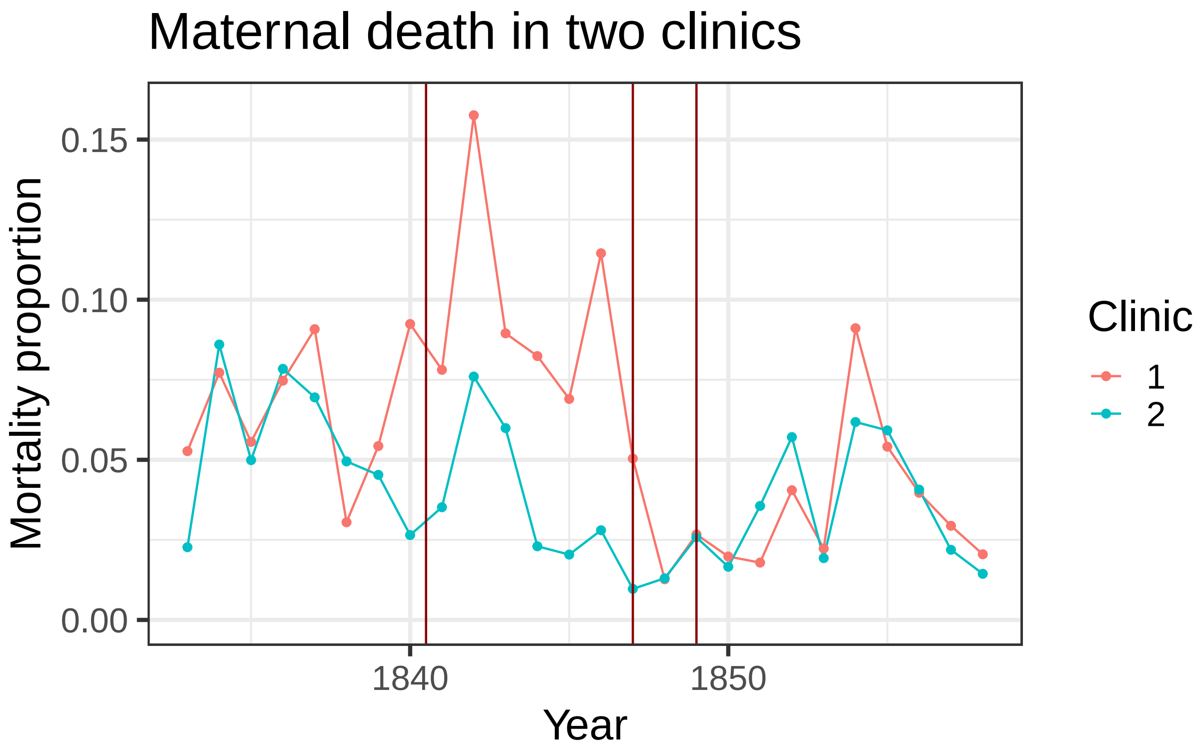


2 Public health

2.1 Maternal mortality

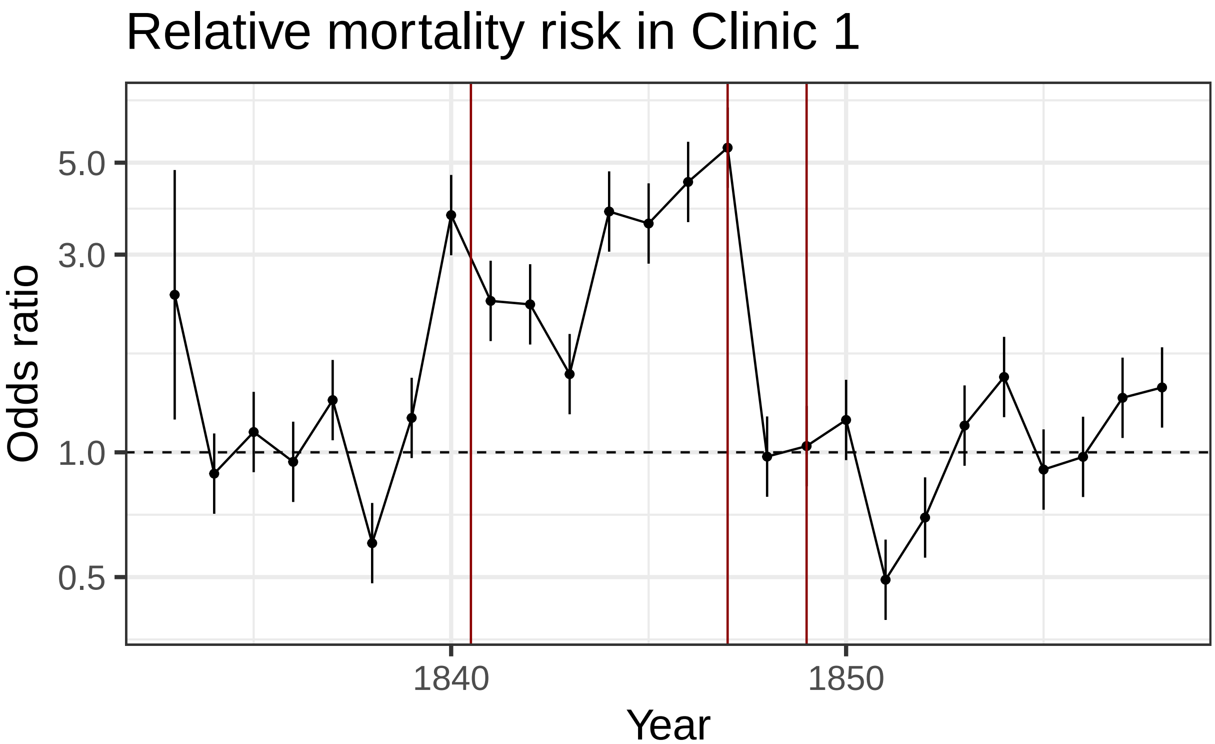
Observation and action

- In 1840, medical students stopped visiting Clinic 1
- In 1847, a surgeon died from infection following a scalpel injury
  - Igor Semmelweiss made medical students wash their hands



Looking at the data

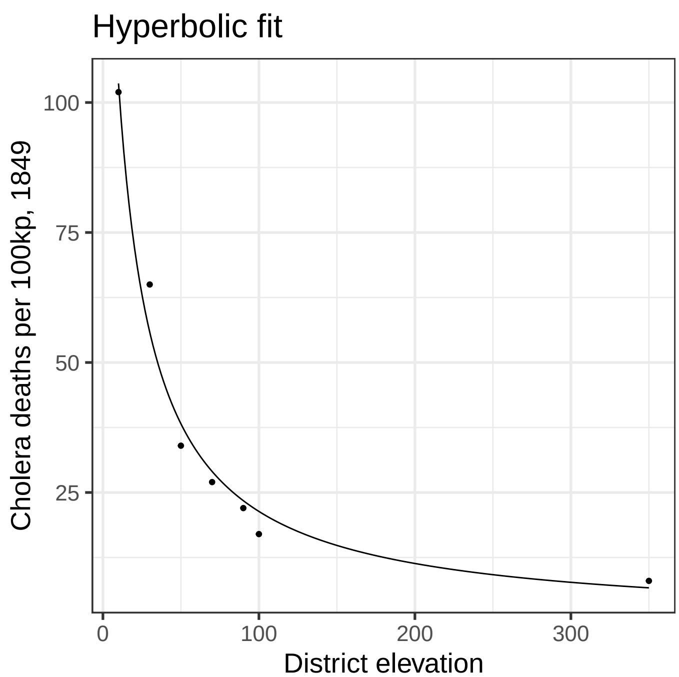
-
-
-
-



2.2 Cholera

- Is it caused by bad air, or bad water?
- What's bad about it?

Cholera and air



2.3 Yellow fever and malaria

- Ross determined the cause of malaria primarily by experiments on mosquitoes
- Reed determined the cause of yellow fever primarily by experiments on human volunteers

3 Approaches to epidemiology

Data, models and science

- We're never finished, we compare models to data over and over again
- Data is what we use to develop and understand models
- Models are what we use to interpret data
  - and they can suggest what data we need to collect
- Complicated or hard-to-test theories may require *dynamical* models

Classical epidemiology

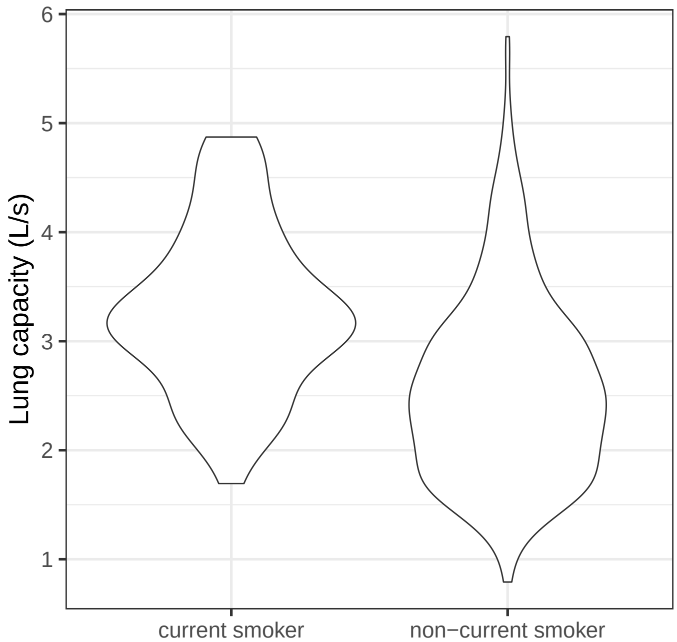
- Avoid mechanism
- Control for non-independence of "units"

Dynamical epidemiology

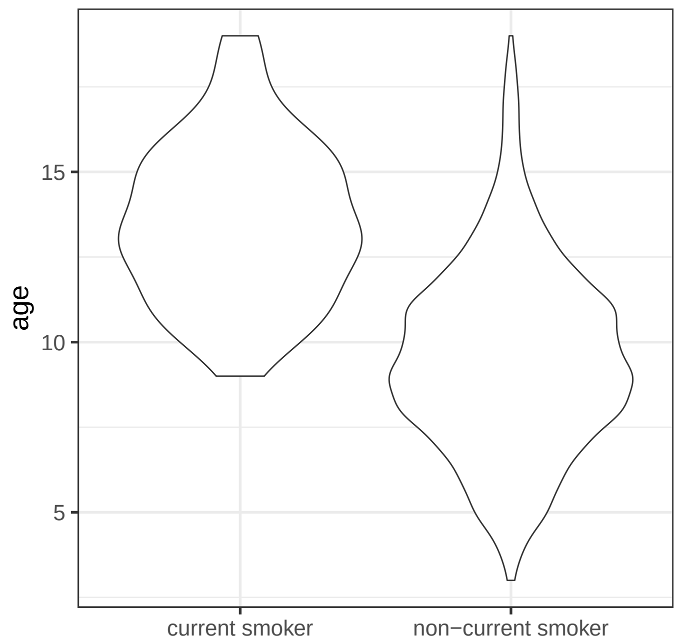
- Embrace mechanism
- Explicitly incorporate dependence between units
    - X is infected because Y infected them

3.1 Classical epidemiology

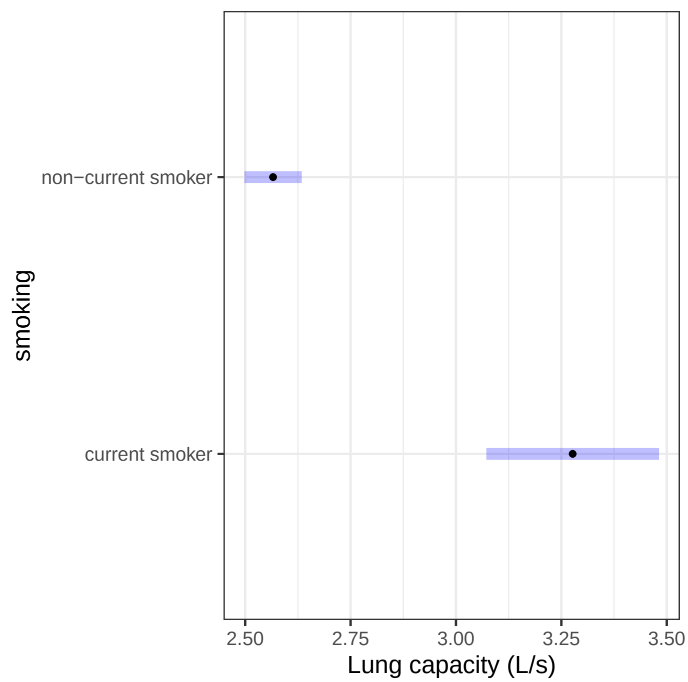
Classical example



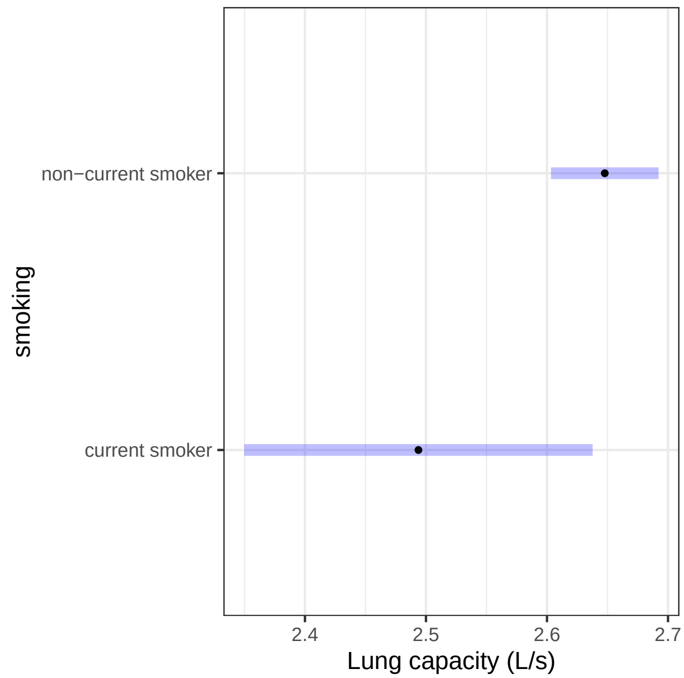
Classical example



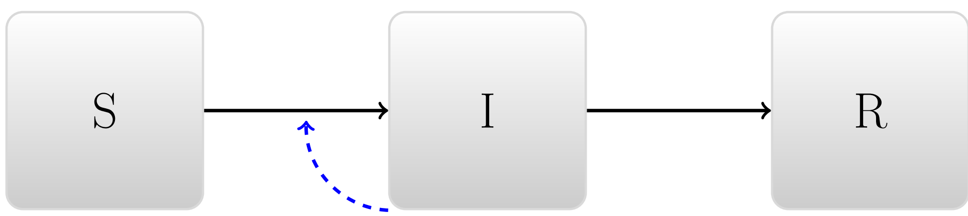
Univariate means



Multivariate means



3.2 Dynamical epidemiology



Other viruses

Pictures from CDC Pink book `https://www.cdc.gov/vaccines/pubs/pinkbook/index.html`

- Rubella
- Measles
- Polio
- Influenza

Bridging

- Classical epidemiology relies on statistics, avoids mechanism
- Mathematical epidemiology (the traditional approach to dynamical epidemiology) explores mechanism, avoids statistics
- Much modern dynamical epidemiology seeks ways to put dynamical mechanisms into a statistical framework
  - This is hard

3.3 Building knowledge from data

- We must have communication up and down the analysis pipeline
  - Data are collected in the field
  - Organized and documented
  - Protected (for confidentiality, and often for the rights of collectors)
  - Summarized
  - Modeled

Example: COVID incidence

- Positive test results but no negative test results
  - Can't correct for testing intensity
- Positive and negative test results, but no individual identifiers
  - Can't correct for multiple testing of the same people
- Test results, but not reason for testing
  - Can't correct for testing focus

Example: COVID variants

- Mutational screens not linked to invididuals
  - We can estimate mutations, but not variants
- Reasons for screening or sequencing not provided
  - Can't correct for selection bias
- Personal information not provided
  - Can't look for geographical patterns, vaccine effectiveness, . . .

Example: West Africa Ebola Outbreak

- Medical vs. public-health priorities
  - Am I responsible if my data request increases the pressure on a front-line responder?
- Individual-level vs. population considerations
  - When it is OK to randomize people to receive a placebo vaccine or treatment?

Example: Microbiome studies

- In some fields, the amount of apparently high-quality data is far outstripping the ability to understand it
- Lots of reasonably intelligent experiments (or observational designs)
- Huge lists of taxonomic (or metagenomic) communities

https://www.nature.com/articles/s41598-021-92400-y

Data vs. models

- Models can teach us a lot, but good data with a simplistic model is usually better than poor (or poorly contextualized) data with a good model
- Sometimes, the most valuable thing about the model is that it helps us figure out what data we need
  -

4 Summary

- Science is an ongoing process
- Models are the way that we bridge between theory and reality
- We can only bridge to reality if we can measure reality
  - Collect and curate good data
- Dynamical models have a key role
  - When we can't do experiments
  - When mechanisms are complex
- We should work to combine dynamics with statistical approaches