Data, models and science

Jonathan Dushoff, McMaster University

MMED 2021

Outline

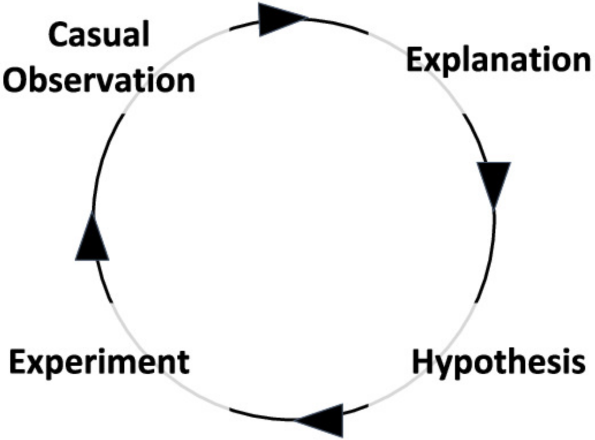
Goals

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

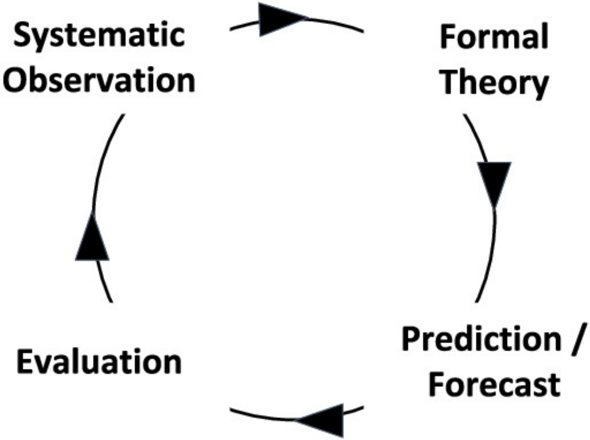
Science is a *process*

- 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

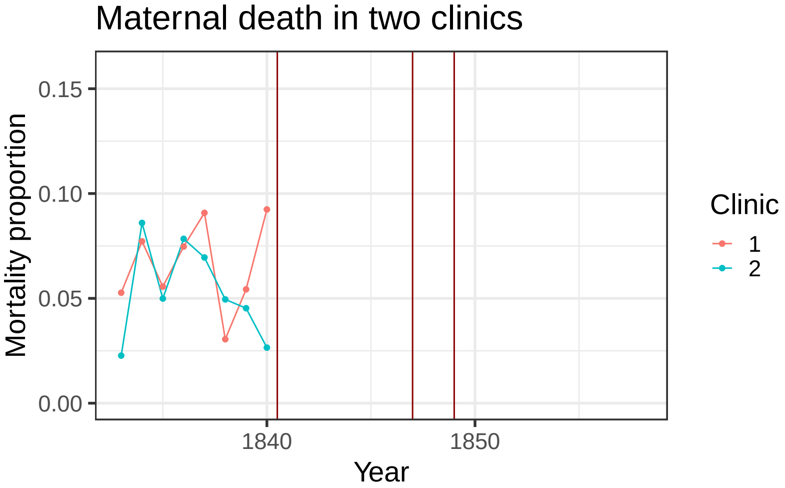


Science without experiments



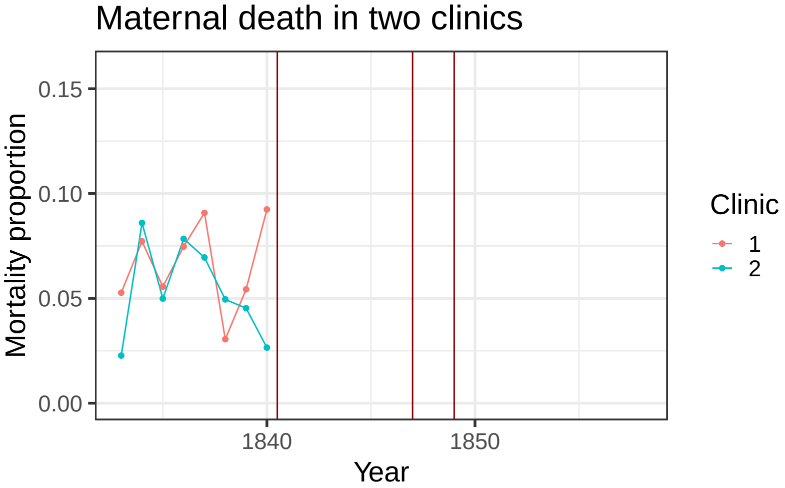
Outline

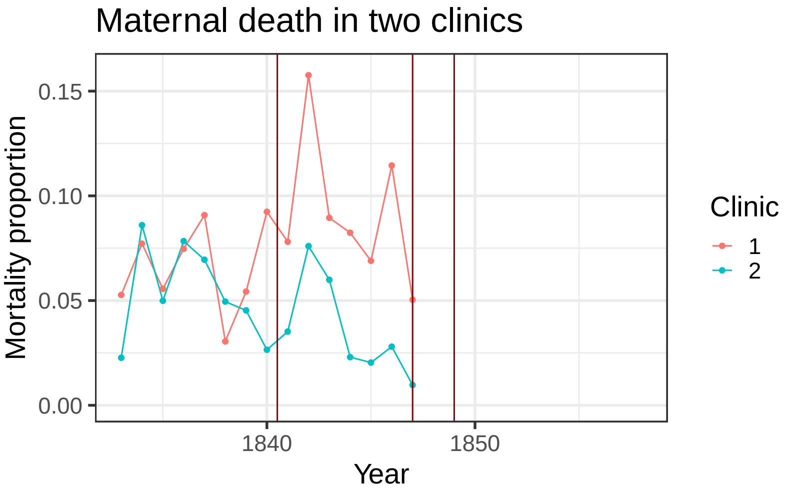
Outline


Maternal death in two clinics
Mortality proportion
0.15
0.10
0.05
0.00
1840
1850
Year
Clinic
1
2

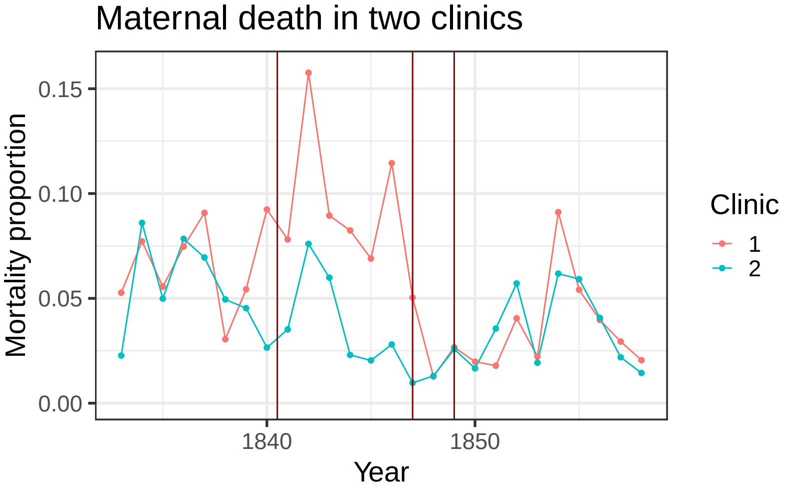
Observation and action

- 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


Maternal death in two clinics
0.15
0.10
0.05
0.00
Mortality proportion
1840
1850
Year
Clinic
1
2



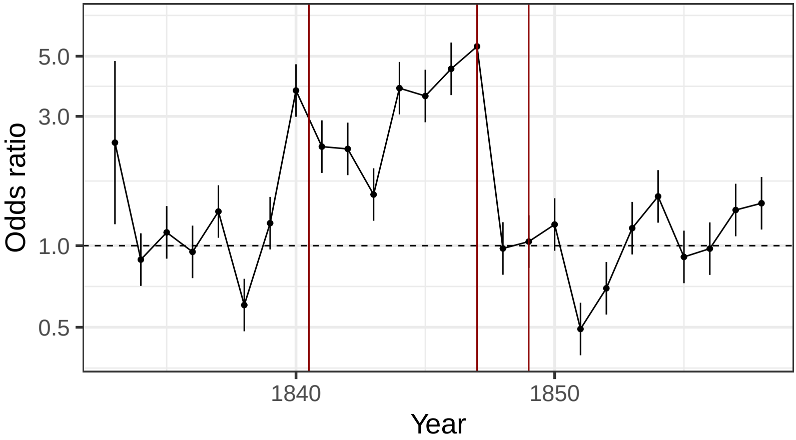
Maternal death in two clinics
0.15
0.10
0.05
0.00
Mortality proportion
1840
1850
Year
Clinic
1
2


Maternal death in two clinics
Mortality proportion
0.15
0.10
0.05
0.00
1840
1850
Year
Clinic
1
2

Looking at the data

- * Why was Clinic 1 so dangerous in the 40s?
- * … and so safe in 1851 and 1852?
- * What can we learn from modern statistics?
- * And what can't we learn without more data?

Relative mortality risk in Clinic 1

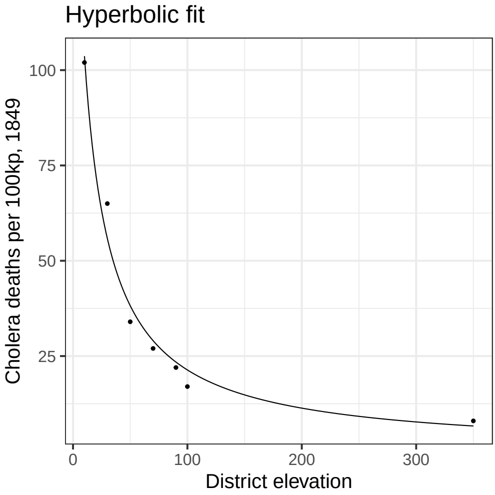


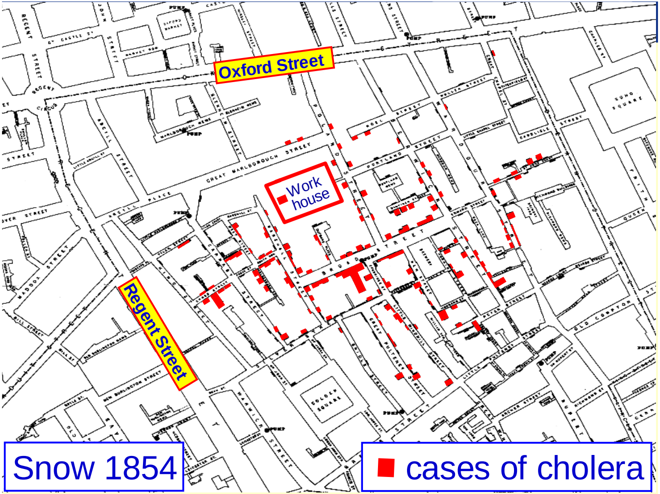
Outline

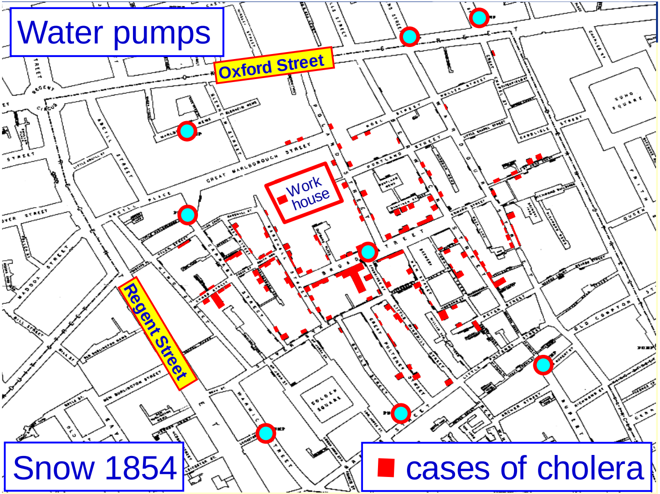
Cholera

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

Cholera and air




Oxford Street
Work house
Regent Street
Snow 1854
cases of cholera


Water pumps
Oxford Street
Work house
Regent Street
Snow 1854
cases of cholera

Outline

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

Outline

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

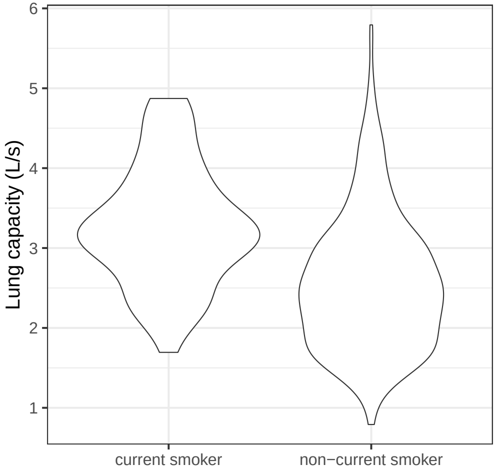
- Avoid mechanism
- Control for non-independence of “units”

Dynamical epidemiology

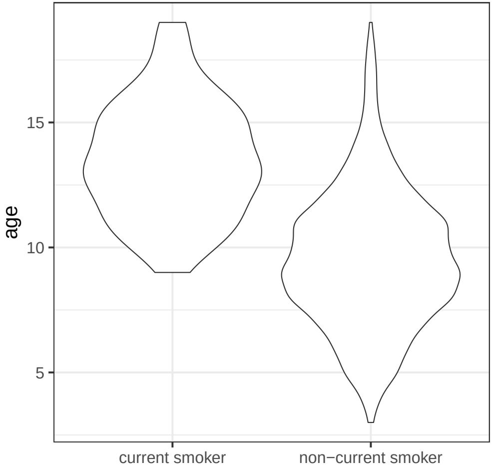
- Embrace mechanism
- Explicitly incorporate dependence between units
  - X is infected because Y infected them

Outline

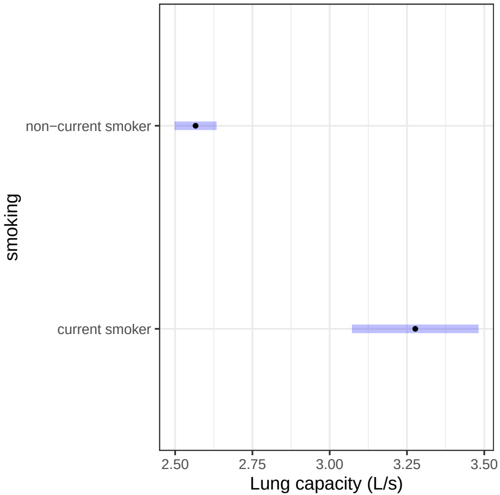
Classical example



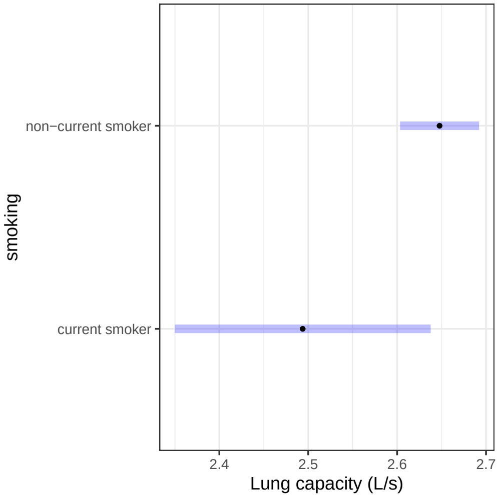
Classical example



Univariate means

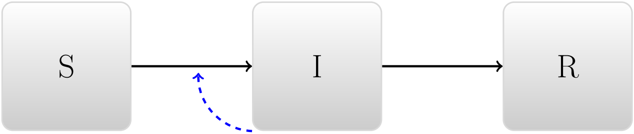


Multivariate means



Outline

Dynamical epidemiology



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

Outline

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

scientific reports

Explore content ⌄ Journal information ⌄ Publish with us ⌄

nature > scientific reports > articles > article

Article | Open Access | Published: 21 June 2021

SARS-CoV-2 viral load in nasopharyngeal swabs is not an independent predictor of unfavorable outcome

Sonsoles Salto-Alejandre, Judith Berastegui-Cabrera, Pedro Camacho-Martínez, Carmen Infante-Domínguez, Marta Carretero-Ledesma, Juan Carlos Crespo-Rivas, Eduardo Márquez, José Manuel Lomas, Claudio Bueno, Rosario Amaya, José Antonio Lepe, José Miguel Cisneros, Jerónimo Pachón ✉, Elisa Cordero, Javier Sánchez-Céspedes & The Virgen del Rocío Hospital COVID-19 Working Team

Scientific Reports **11**, Article number: 12931 (2021) | Cite this article

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
  - * Value-of-information models

Outline

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

Title: Data, models and science

Attribution: Jonathan Dushoff, McMaster University, MMED 2021

Source URL: http://www.ici3d.org/MMED/schedule/

For further information please contact admin@ici3d.org.