MMED
African Institute for the Mathematical Sciences
Muizenberg, South Africa
June, 2015

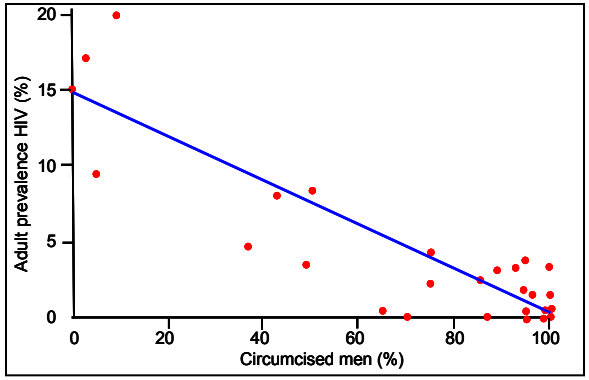
Thinking About Data

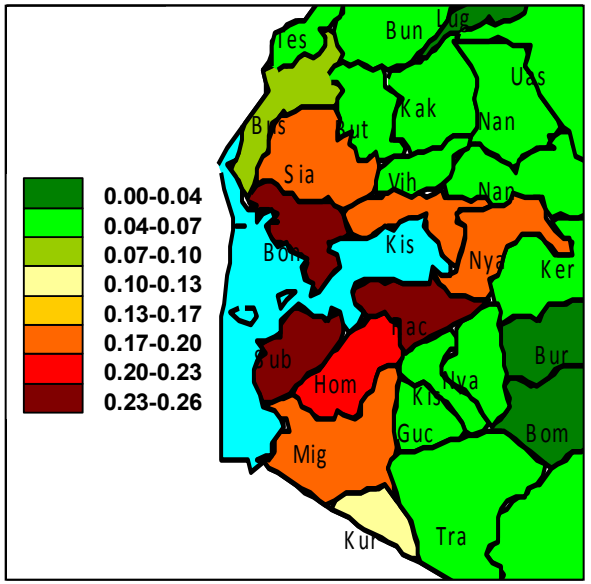
Wim Delva, MD, Pd.D.
Brian Williams, Ph.D.
Jim Scott, Ph.D, M.A., M.P.H.

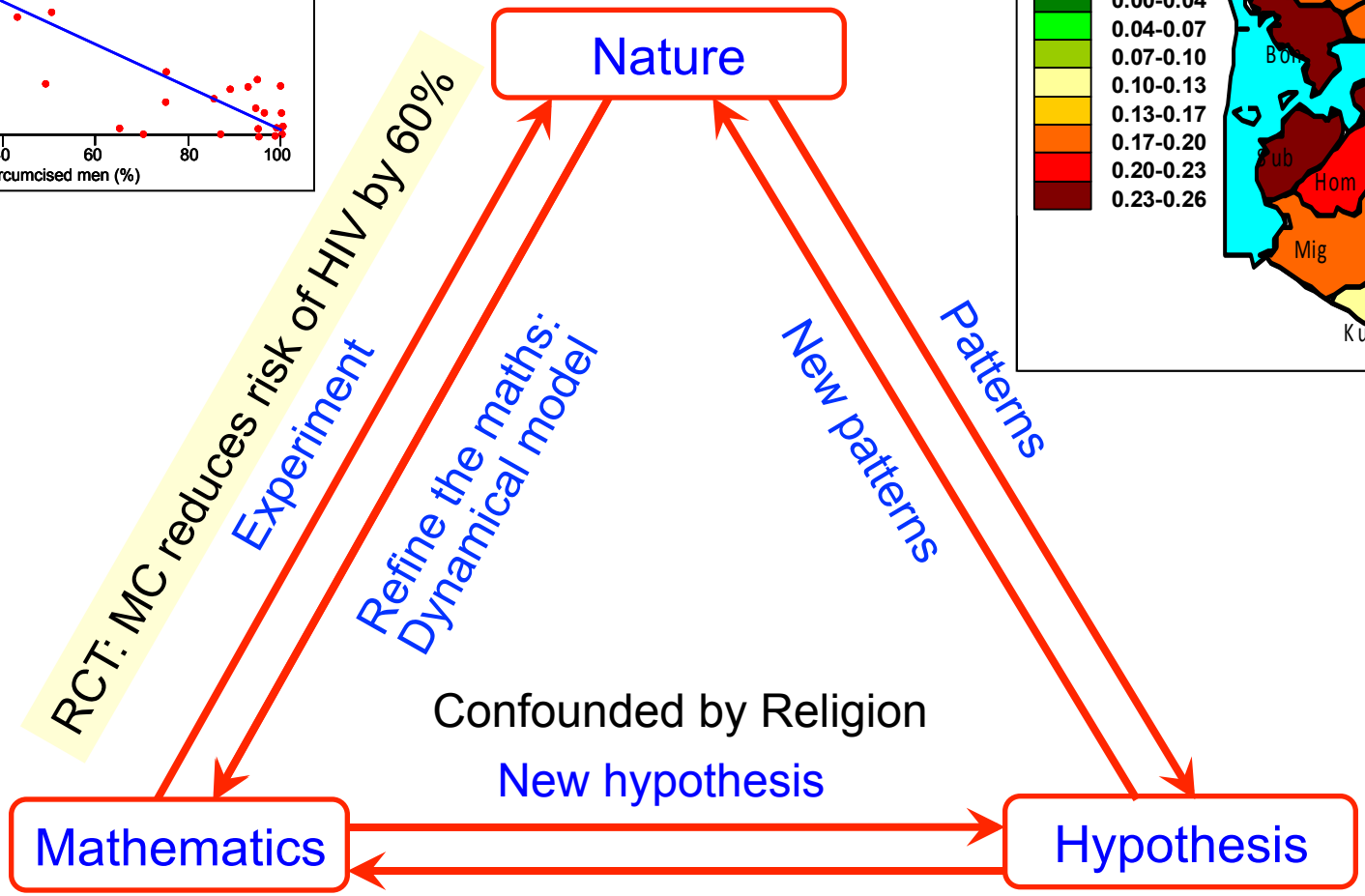
The object of all science, whether natural science or psychology, is to co-ordinate our experiences into a logical system.

Einstein A. *The Meaning of Relativity* (1922)

What is Science





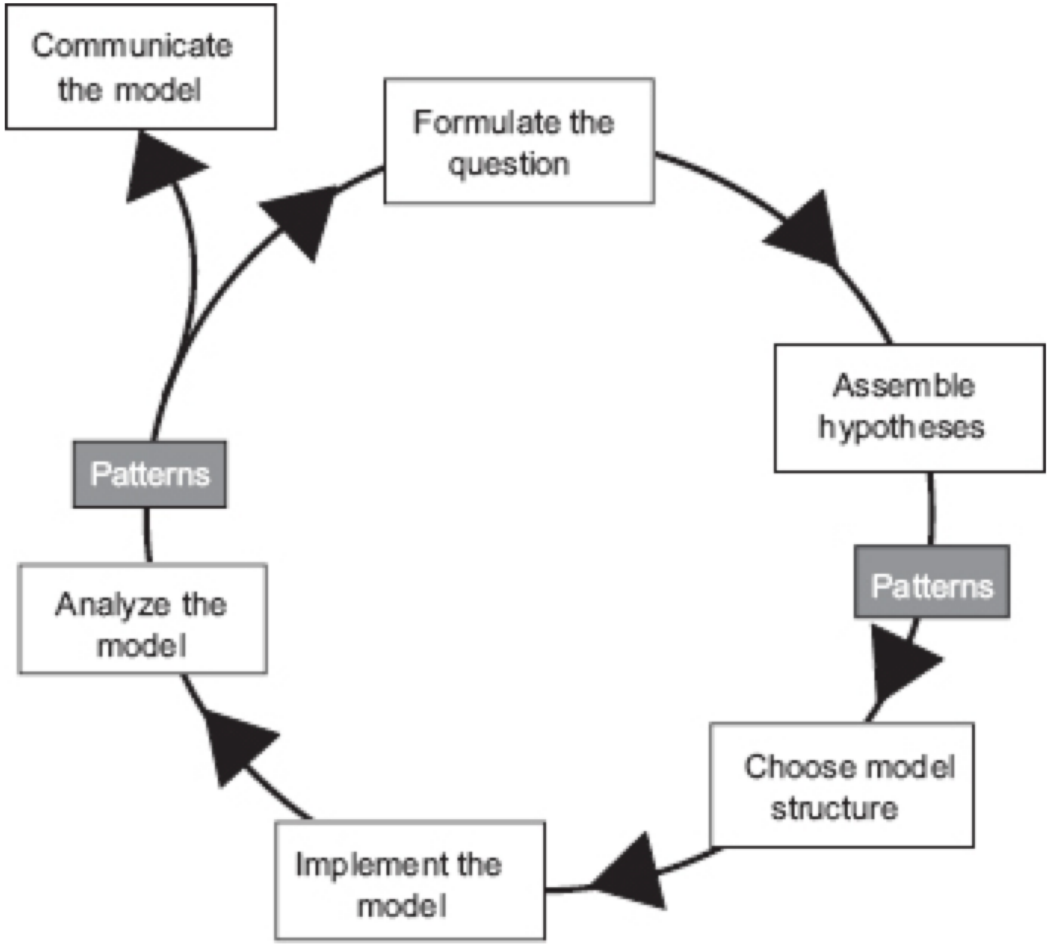


Male circumcision accounts for the variation

Include religion in the model

Modern science is the mathematization of hypothesis about nature combined with rigorous expermentation—Joseph Needham

The modelling cycle



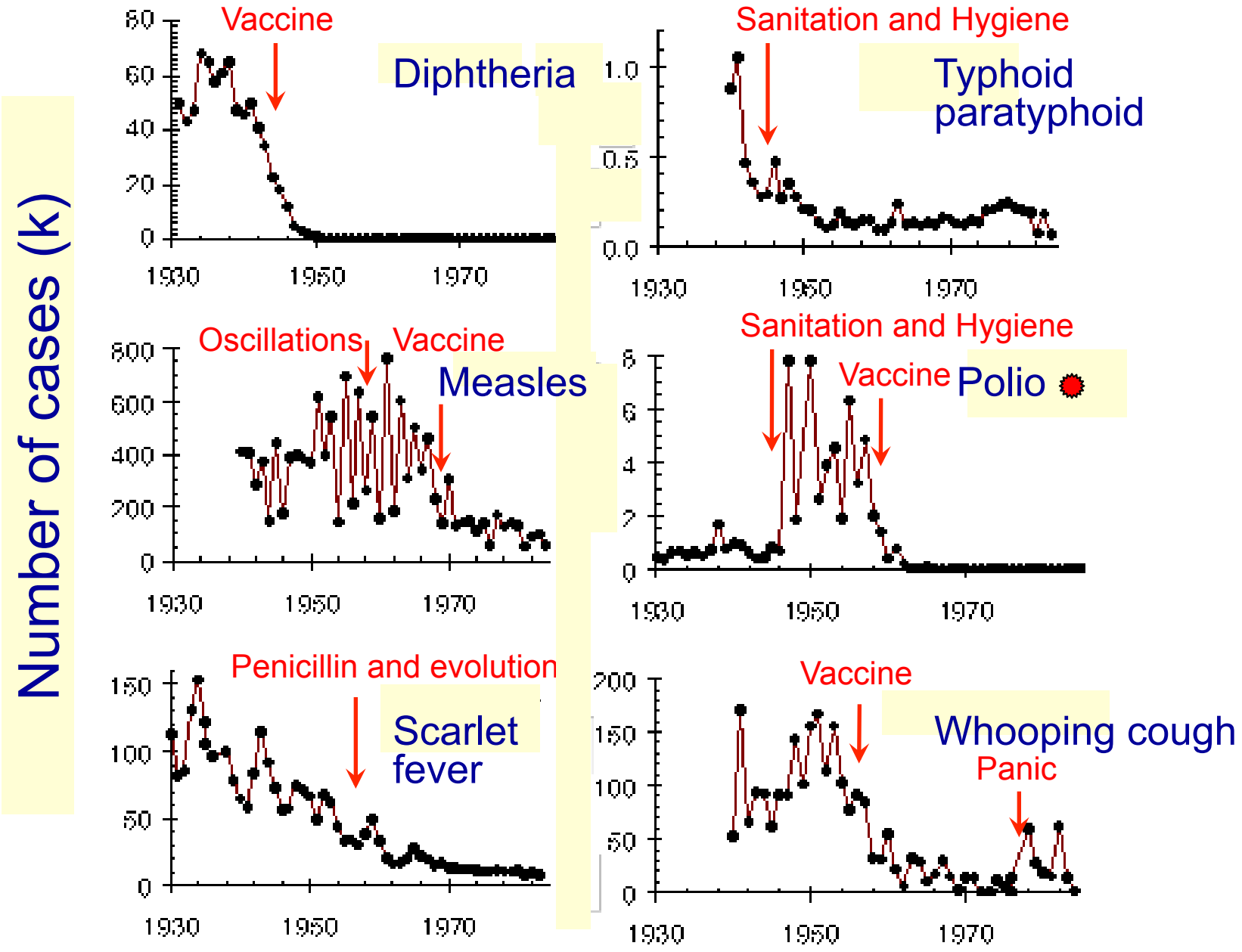
… the supreme goal of all theory is to make the … basic elements as simple and as few as possible without … surrendering … a single datum of experience.

—Albert Einstein, 1933

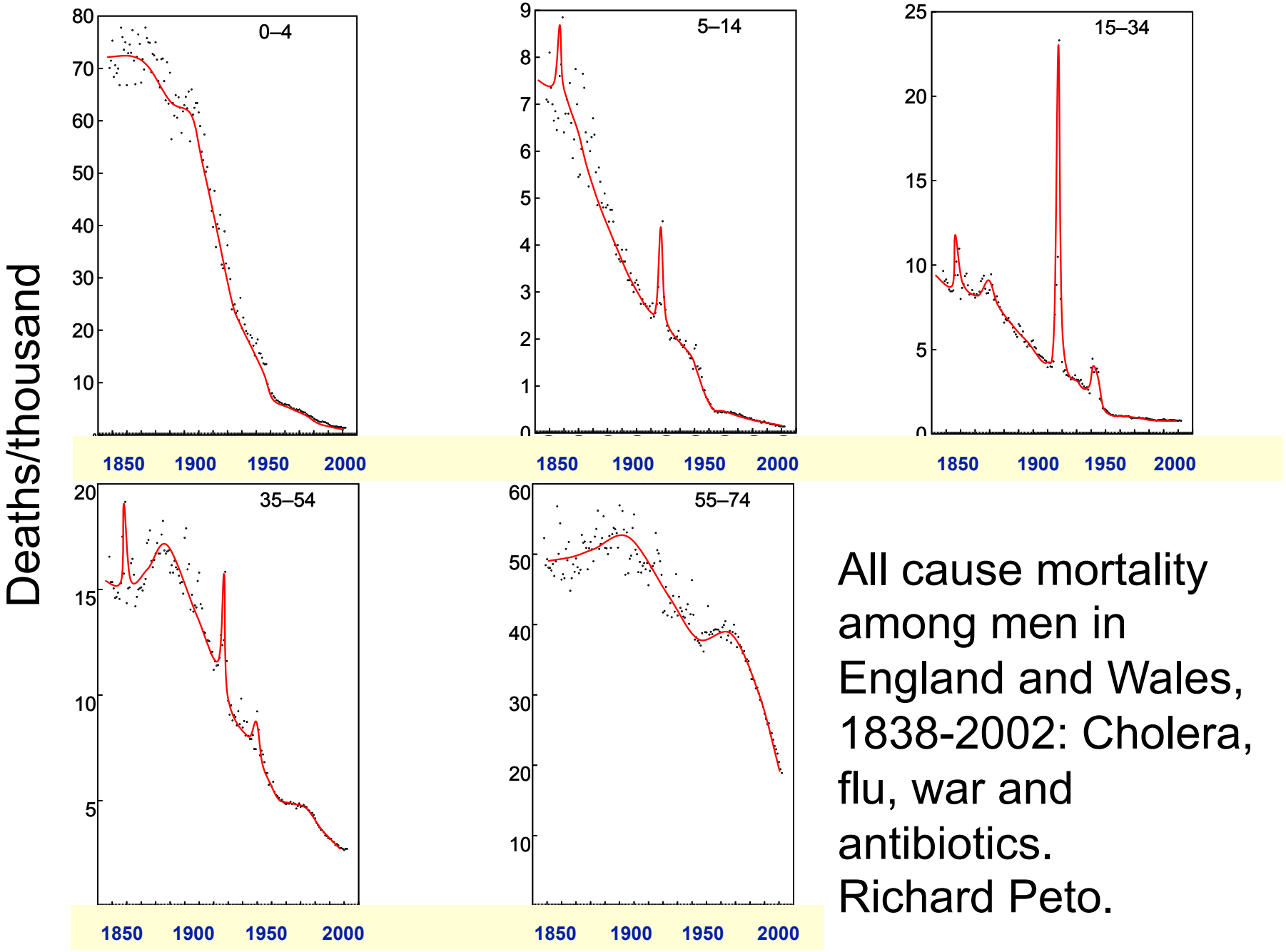
… the supreme goal of all theory is to make the … basic elements as simple and as few as possible without … surrendering … a single datum of experience.

—Albert Einstein, 1933

Only introduce complexity as the data or the question demand it.



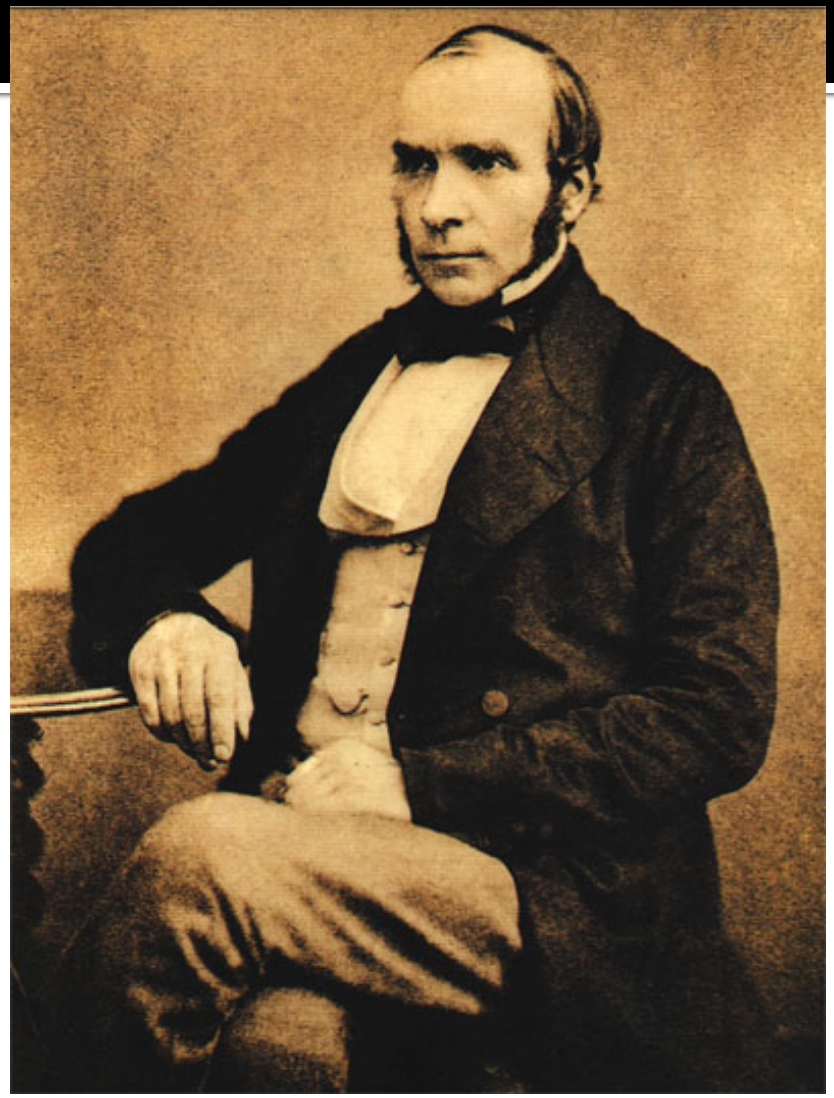
Diseases in England and Wales, 1930 to 1980



All cause mortality among men in England and Wales, 1838-2002: Cholera, flu, war and antibiotics. Richard Peto.

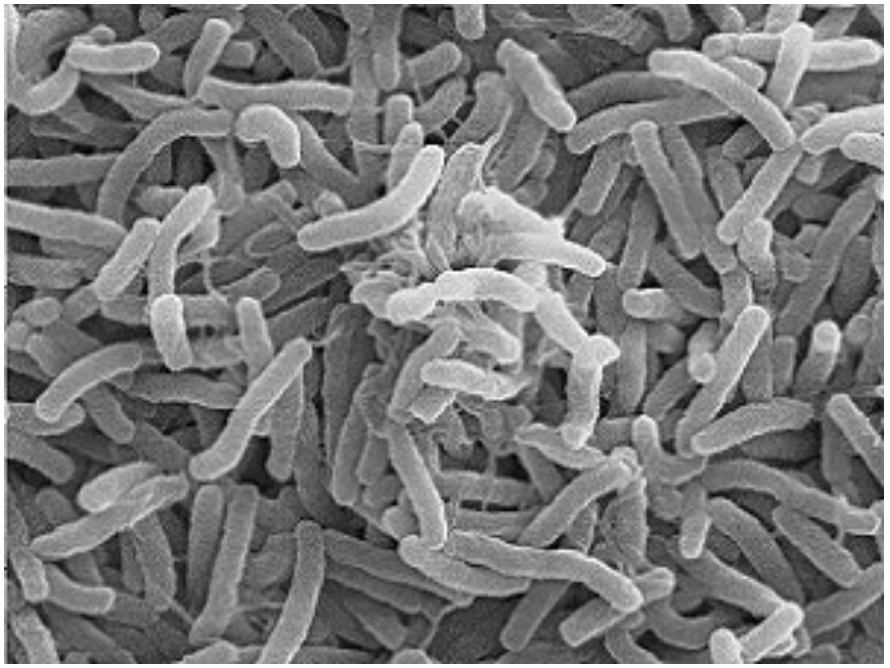
John Snow

- Considered father of field epidemiology
- Conducted a series of investigations of cholera outbreaks in London
- Acclaimed Anesthesiologist



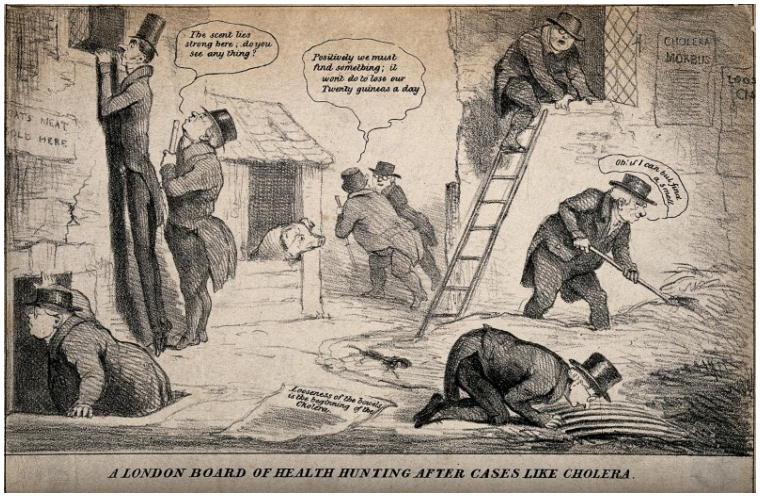
London Cholera Epidemics

- 4 outbreaks between 1831 and 1854
- Most believed cholera was transmitted through vapors – "Miasma"
- Snow proposed that cholera was actually spread by contaminated water
- Considered "folly"



Snow: Observations of Source Water

- Snow had noted that the two water supply companies, Lambeth Co. and Southwark and Vauxhall Co., were drawing water from Thames at a point downstream of London
- This water was heavily polluted – a likely source of infection



Snow: "The Grand Experiment"

- By 1854, the Lambeth Company had moved their water intake point upstream of London
- Snow decided to compare deaths from cholera in July-August 1854 among households served by the two different companies
  - Walked door-to-door to determine the water source for each house

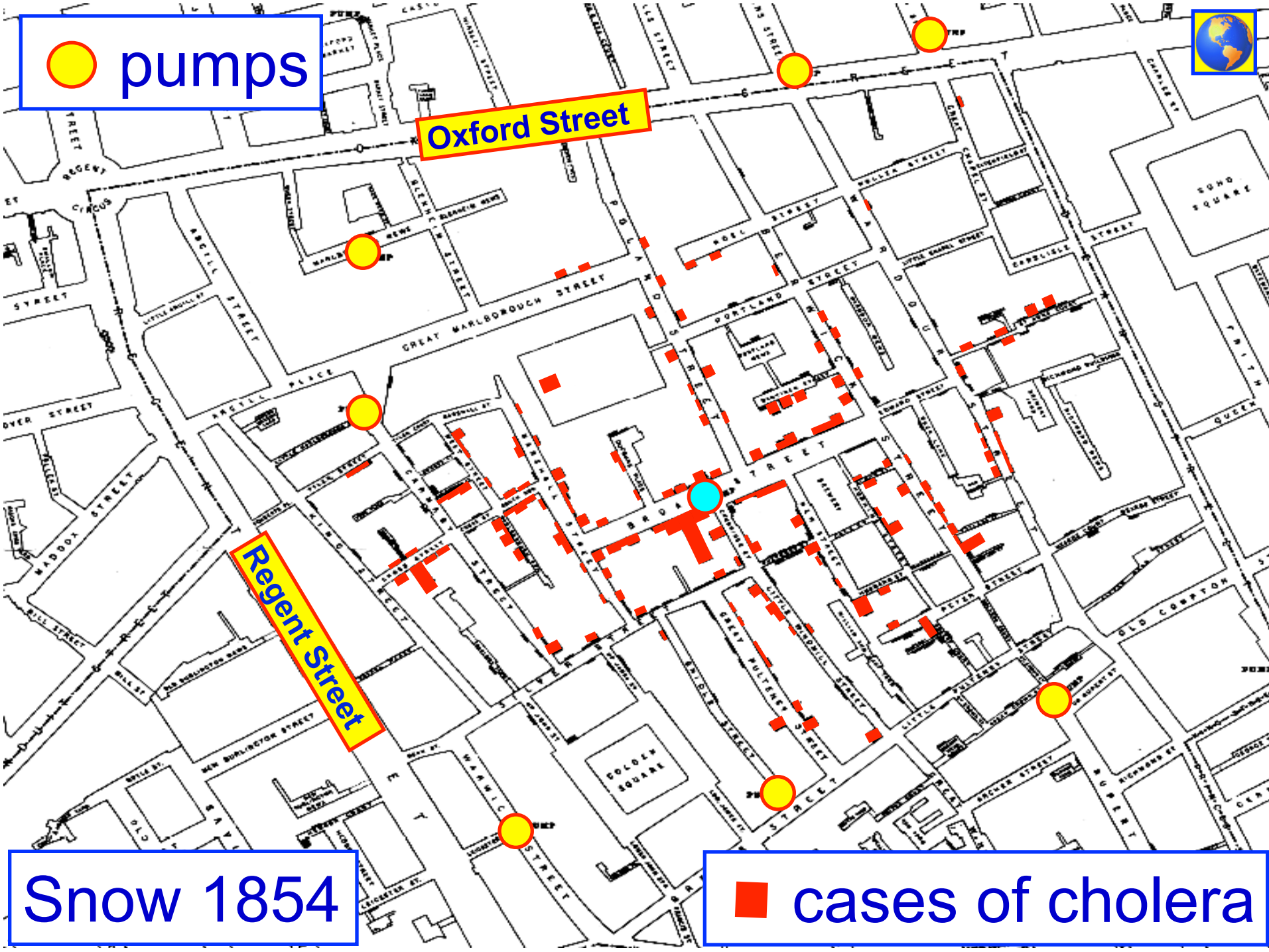
Snow: Comparison of Mortality Rates

- July 9th to August 26th, 1854
- Mortality much higher in households that were supplied with drinking water from the Southwark and Vauxhall Company

Districts with Water Supplied by	Population (1851 Census)	Deaths from Cholera	Cholera Death Rate per 1,000 Population
Southwark and Vauxhall Co. only	167,654	844	5.0
Lambeth Co. only	19,133	18	0.9
Both companies	300,149	652	2.2

Snow: Epidemic in Golden Square

- In addition to noting role of the water companies, Snow also described the epidemic that hit the Golden Square area of London
- He determined where cholera cases lived and worked, and created a map

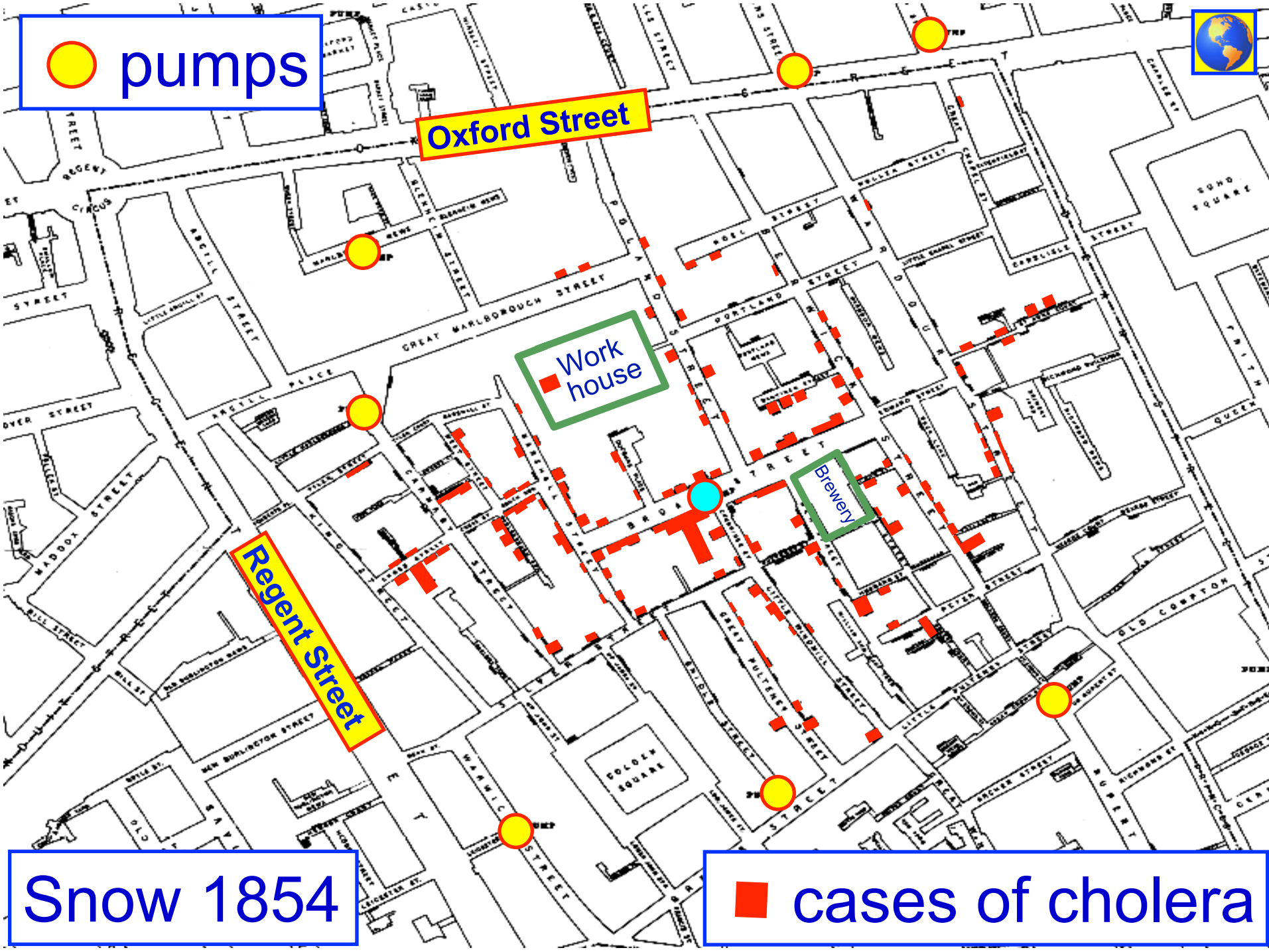

pumps
Oxford Street
Regent Street
Snow 1854
cases of cholera

Snow: Hypothesis generation

- Since Snow believed water was the source of infection, he also plotted water pumps on map
- More of the cases were clustered around the Broad Street pump, than around other pumps

Snow: Investigating Anomalies in the Data

- Although these findings provided more support for his theory, he noted that there were relatively few cases in 2 of the blocks near the pump
- He conducted more field epidemiology to characterize that neighborhood


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

Snow: Factors protective against disease

- The work house
  - A large work house in the area had a deep well that served as the only source of drinking water
- The brewery
  - A nearby brewery also supplied its workers with a daily allotment of beer

Snow: Public Health Action

- Snow took these findings to the appropriate health authorities and, as the story goes, removed the handle from the Broad Street pump
- Shortly afterwards, the cholera epidemic subsided

“Why is it, then, that Dr. Snow is so singular in his opinion? Has he any facts to show in proof? No……The fact is, that the well whence Dr. Snow draws all sanitary truth is the main sewer. In riding his hobby very hard, he has fallen down through a gully-hole and has never since been able to get out again”

- The Lancet (editorial)

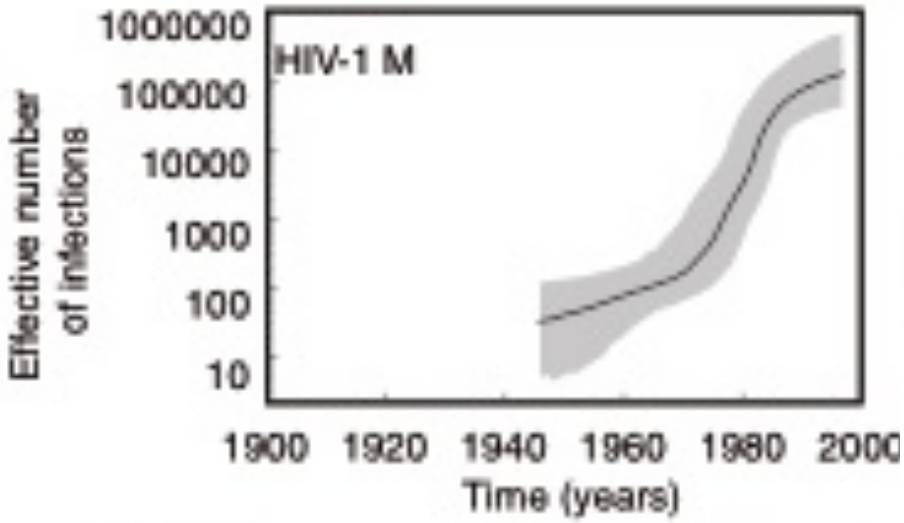
source: The Ghost Map, by Stephen Johnson 2006

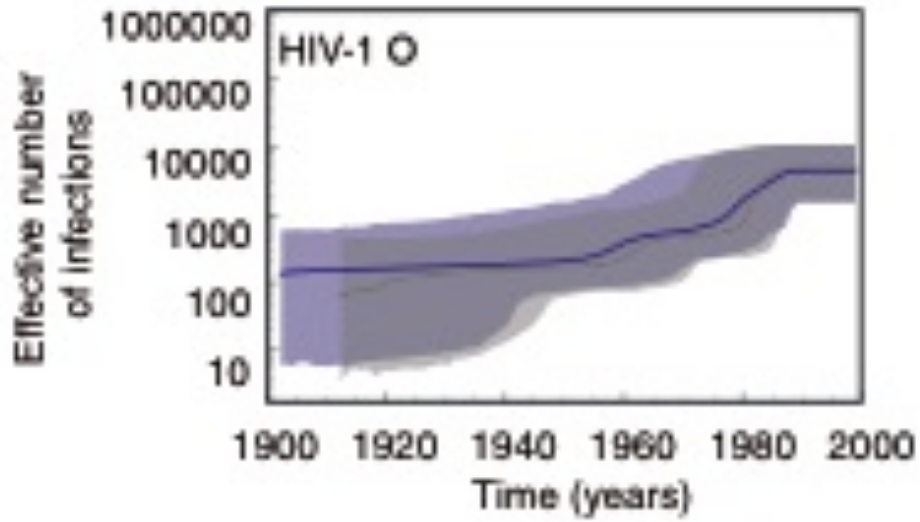
Other data-driven investigations

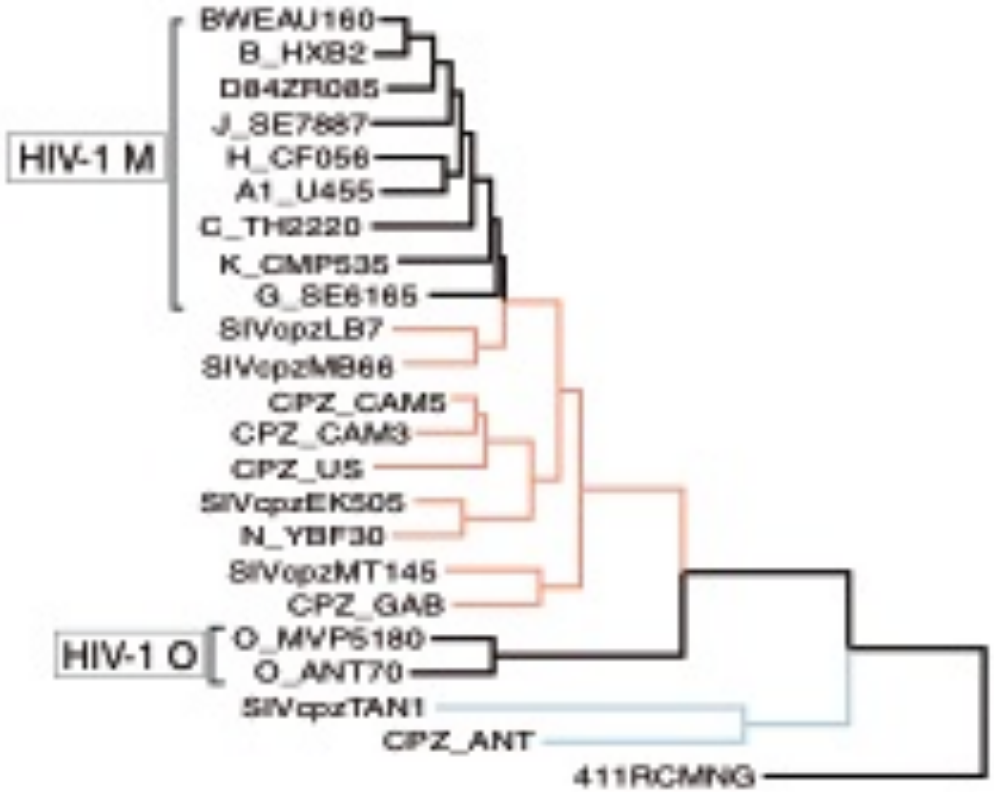
- The origin of HIV:
  - Contaminated Polio vaccines?
  - Pamaquine malaria prophylaxis trial gone wrong?
  - Zoonosis?

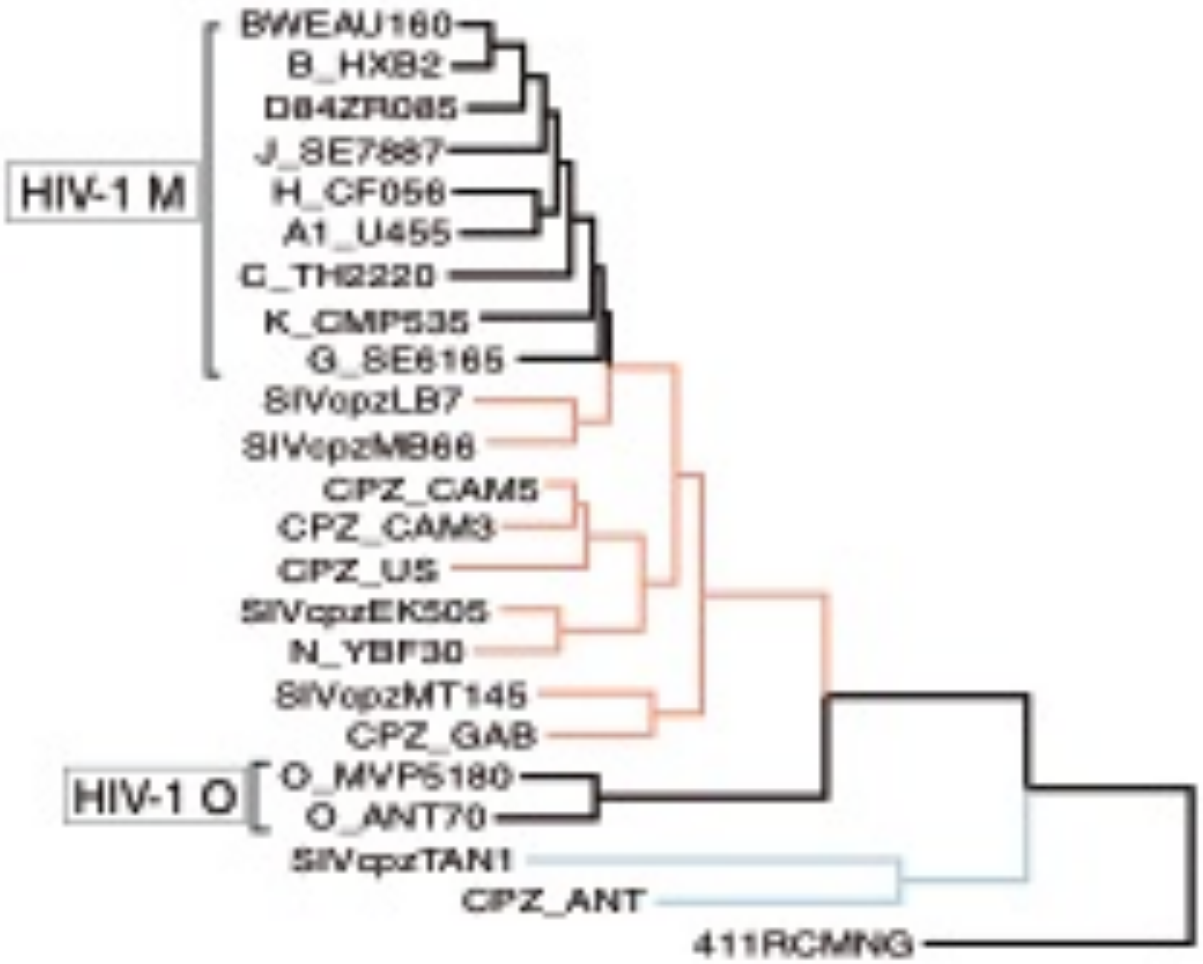
  - Demography – Social science – Molecular biology – anthropology

The origin of HIV

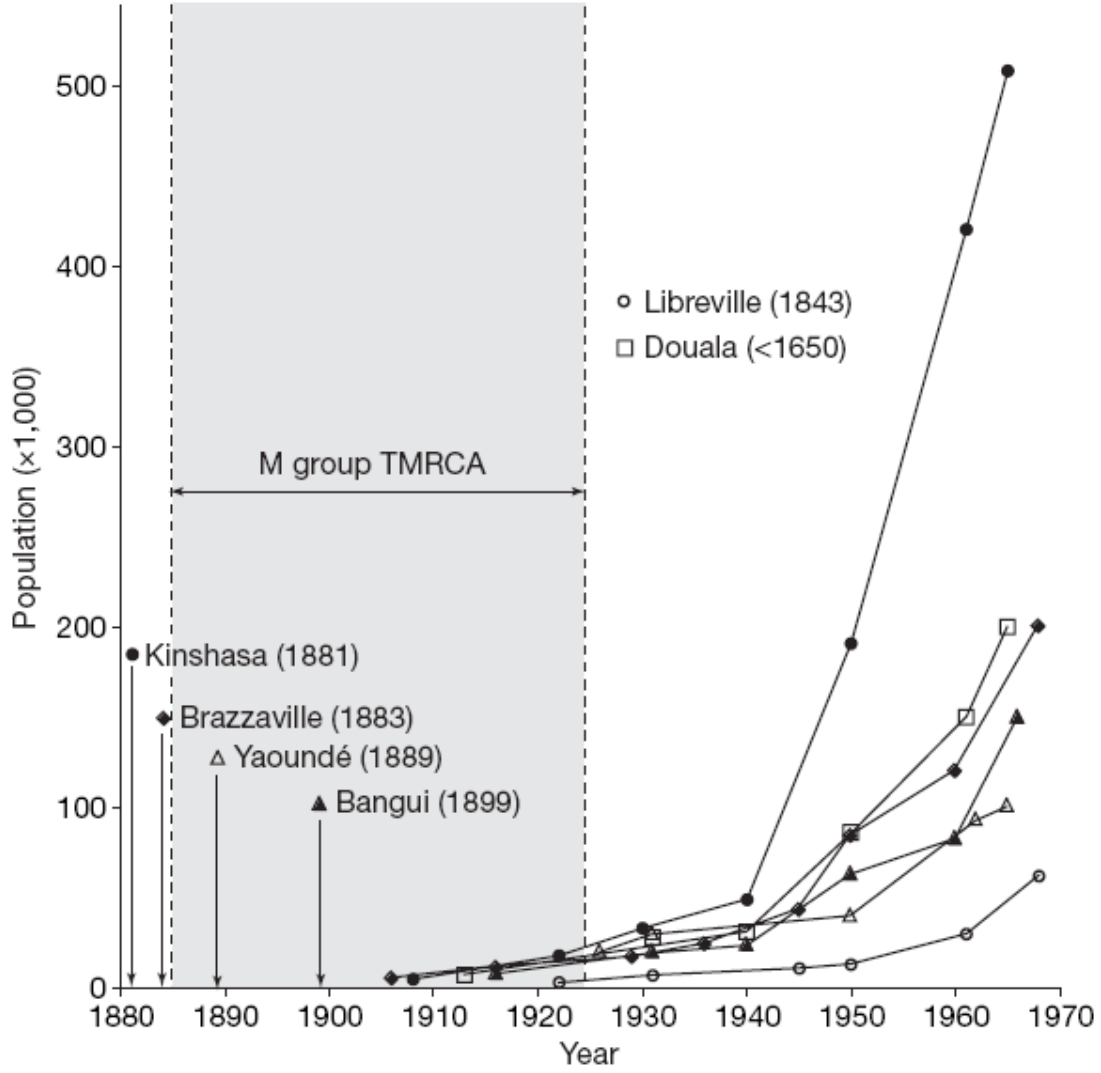







HIV-1 M
BWEAU160
B_HXB2
D84ZR085
J_SE7887
H_CF056
A1_U455
C_TH2220
K_CMP535
G_SE6165
SIVcpzLB7
SIVcpzMB66
CPZ_CAM5
CPZ_CAM3
CPZ_US
SIVcpzEK505
N_YBF30
SIVcpzMT145
CPZ_GAB
HIV-1 O
O_MVP5180
O_ANT70
SIVcpzTAN1
CPZ_ANT
411RCMNG

The origin of HIV



Worobey et al. Nature 2008

Skepticism vs Openness

- "It seems to me what is called for is an exquisite balance between two conflicting needs: the most skeptical scrutiny of all hypotheses that are served up to us and at the same time a great openness to new ideas …
- If you are only skeptical, then no new ideas make it through to you …
- On the other hand, if you are open to the point of gullibility and have not an ounce of skeptical sense in you, then you cannot distinguish the useful ideas from the worthless ones."

- Carl Sagan

Advice to young epidemiologists

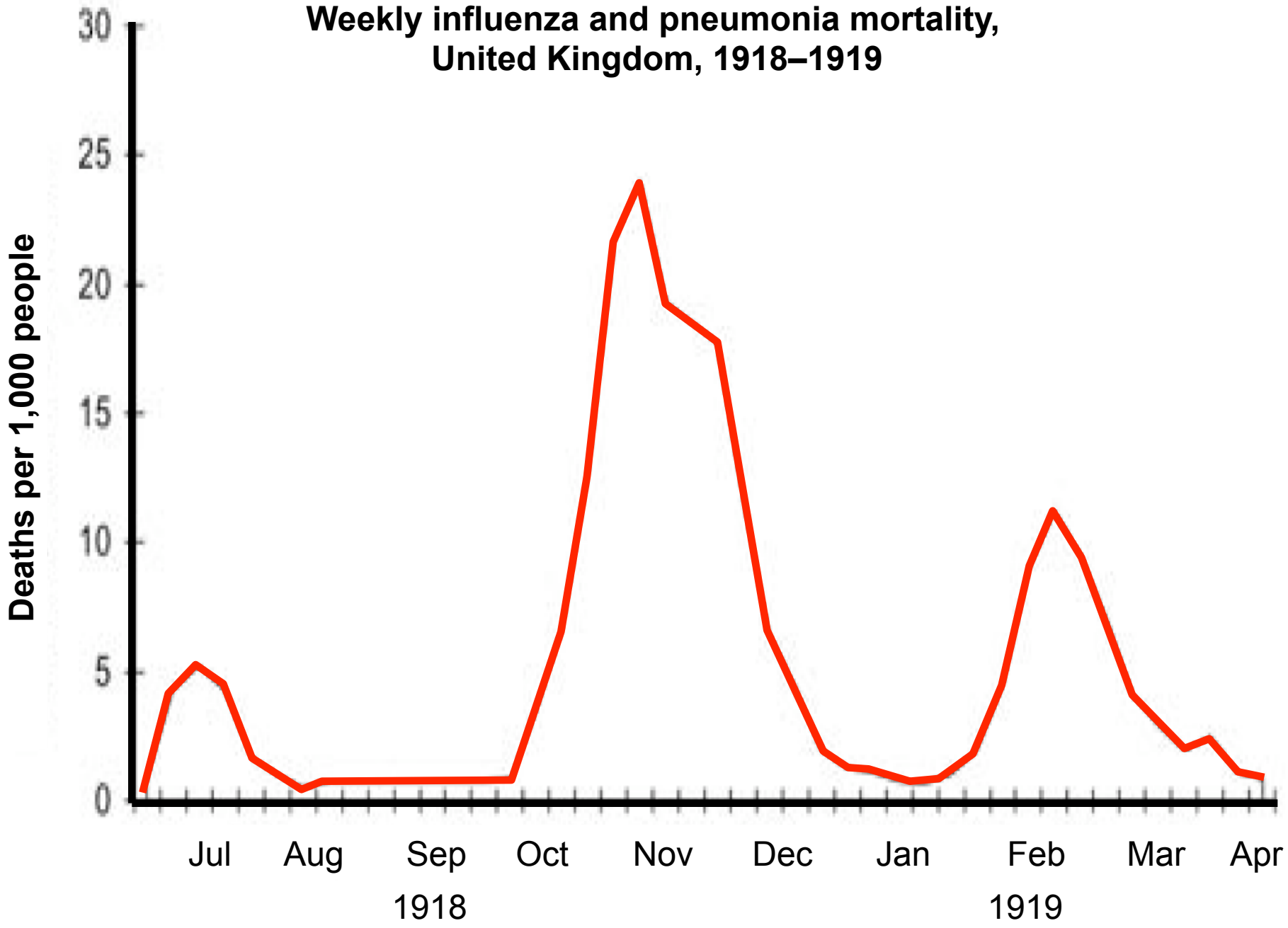
Never make a calculation until you know the answer. Make an estimate before every calculation, try a simple biological argument (R_0, generation time, selection, survival, control). Guess the answer to every puzzle. Courage: no one else needs to know what the guess is. Therefore, make it quickly, by instinct. A right guess reinforces this instinct. A wrong guess brings the refreshment of surprise. In either case, life as an epidemiologist, however long, is more fun.

Plagiarised from E.F. Taylor and J.A. Wheeler *Space-time Physics* 1963

That’s it

Experiment alone can decide on truth but the axiomatic basis of physics cannot be extracted from experiment.

—Albert Einstein, 1933



Taubenberger, J.K. and Morens, D.M. '1918 Influenza: the Mother of All Pandemics' Emerging Infectious Diseases 12 (2006) 15

Snow: Field Epidemiology

- To gain support for his hypothesis, Snow approached residents and cases to determine where they drew their water
- He found out that residents avoided other pumps because they were either grossly contaminated or located too inconveniently for most residents