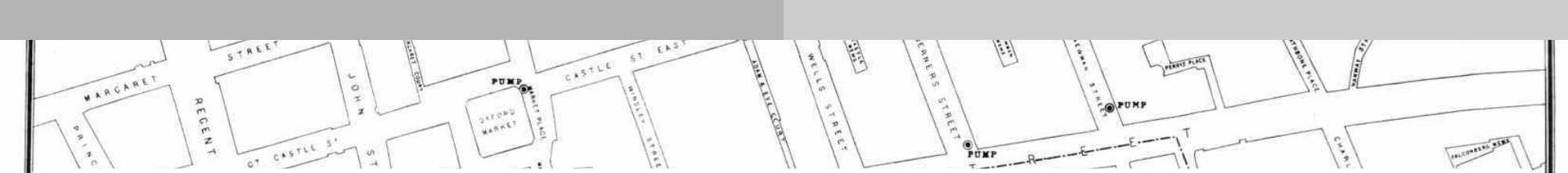


Causality in the Time of Cholera: John Snow as a Prototype for Causal Inference

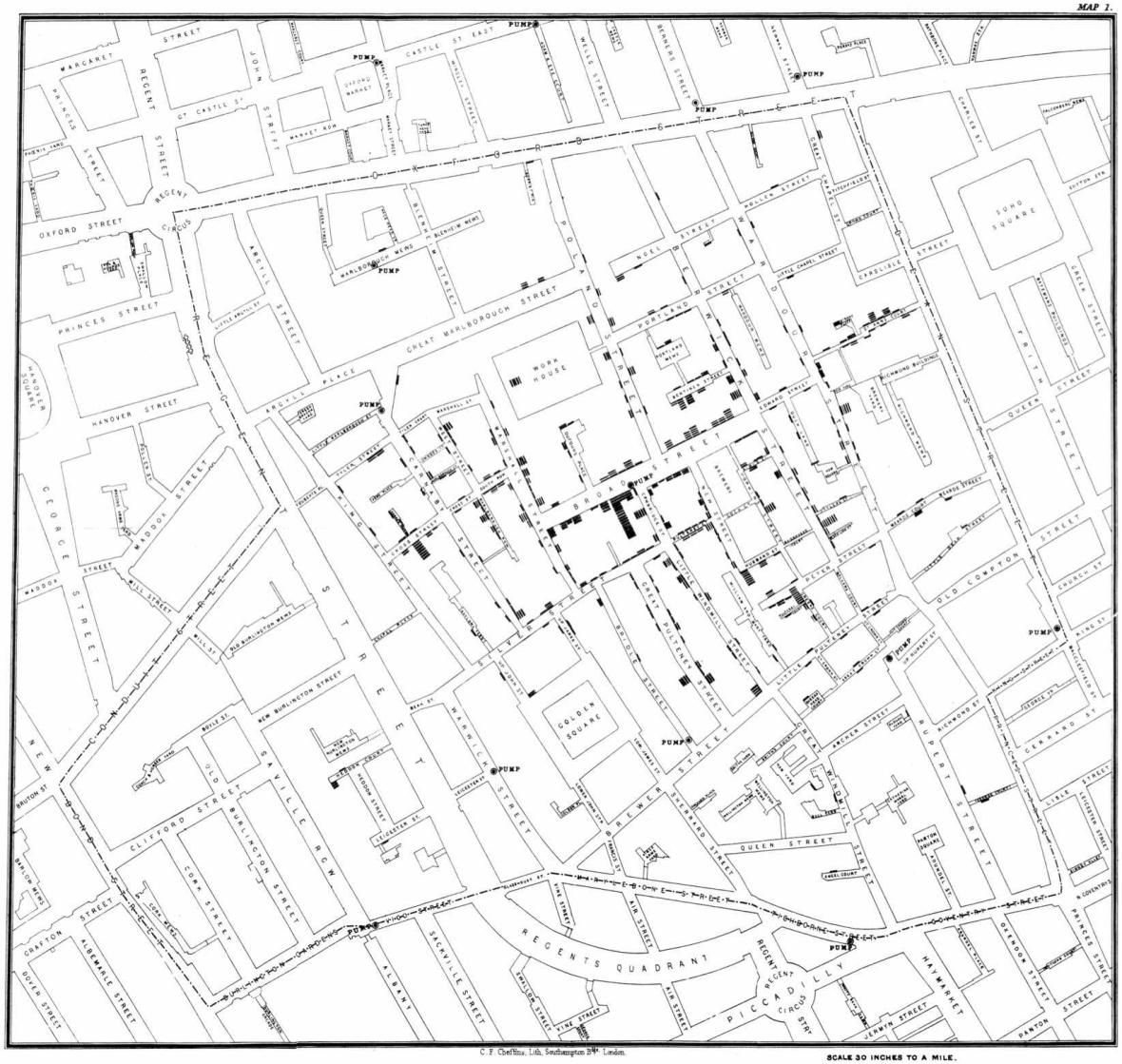
J. Heckman & Economics 312
https://papers.ssrn.com/abstract=3262234

Thomas S. Coleman

Harris (tscoleman@uchicago.edu)

May 7, 2019

John Snow Known for "Broad St Pump" & Mapping



But I'm Studying Snow as a Template for Causal Inference

Snow is fun for three reasons – here focus on (2):

1. **Rollicking Good Tale** – full of heroism, death, and statistics
2. **Causal Inference** – template for how to marshal evidence in support of a causal explanation
3. **Statistics & Instruction** – The data are simple but the analysis demonstrates multiple data analytic tools we use today
   - combining maps and data (GIS or geographic information systems)
   - regression and error analysis
   - difference-in-differences regression
   - natural experiments and randomization

But I'm Studying Snow as a Template for Causal Inference

Snow is fun for three reasons – here focus on (2):

1. **Rollicking Good Tale** – full of heroism, death, and statistics
2. **Causal Inference** – template for how to marshal evidence in support of a causal explanation
3. **Statistics & Instruction** – The data are simple but the analysis demonstrates multiple data analytic tools we use today
   - combining maps and data (GIS or geographic information systems)
   - regression and error analysis
   - difference-in-differences regression
   - natural experiments and randomization

Also humbling reminder: with overwhelming evidence and strong analysis, Snow still failed to convince the medical establishment, the public, or the authorities

Outline

Causal Assessment Procedure – based on Katz and Singer [2007] – preliminary

Katz and Singer propose an "Attribution Assessment Procedure" to weigh the disparate evidence and conflicting explanations associated with reports of a chemical weapons attack. Such an exercise has many similiarities with efforts to determine causal effects in social sciences generally, and cholera in 1850s London in particular. Katz and Singer propose seven steps, which I modify slightly:

> *Possible Chemical & Biological Weapons attacks, 1970s-80s, "Can an Attribution Assessment Be Made for Yellow Rain?"*

Katz & Singer as "Causal Assessment Procedure"

1. Divide evidence into blocks or types of evidence
2. Assign to each block a *veritas* rating – quality of data
3. Develop groups of hypotheses
4. Assess each evidence block for strength of rejection for each hypothesis
   - Consider *rejection* of hypotheses (refute, neutral, consistent) rather than strength of association (support of hypotheses)
5. Organize evidence blocks by hypothesis into matrix
6. Choose hypothesis not contradicted
7. Strongest hypothesis checked

Theory, Data, Hypothesis Testing

Data or Evidence Blocks

Albion Terr

17 houses
single outbreak

Broad St

~10 sq blocks
2wks, 700 deaths

South London

summer/fall 1854
~400k subjects mixed
treated & untreated

Theory & Hypotheses

water & small intestine	miasma (airborne)	elevation, class, ...

Hypothesis or Testing Blocks

Albion Terr

- Narrative

Broad St

- Map
- Cases
- Contin

South London

- Diff-in-Diffs
- Mixing
- No sub-district pop
- With sub-district pop

Cholera – Disease of Poor Sanitation

What is Cholera?

- *Vibrio Cholerae* – bacterium that infects the small intestine of humans
- Causes severe diarrhea (& vomiting) that drains fluids
- Death from dehydration & organ failure
- Oral Rehydration Therapy highly succesfull (roughly 1960s)
  - In case you ever need it, here's the recipe – 1 liter boiled water, 1/2 teaspoon salt, 6 teaspoons sugar, mashed banana (potassium)

Cholera thrives in crowded cities with poor sanitation

- Transmitted through recycling (drinking) sewage
- When cholera exits one victim, needs to find a way into gut of others
- Victorian London was an ideal playground for cholera to thrive

Well-Articulated Theory

> *Most importantly, Snow had a good idea—a causal theory about how the disease spread—that guided the gathering and assessment of evidence. (Tufte)*

Snow proposed his waterborne theory of cholera in the 1849 pamphlet *On the mode of communication of cholera* (Snow [1849]). Without the benefit of the germ theory of disease or any evidence on the bacterium *Vibrio cholerae* Snow nonetheless proposed a consistent (and correct) theory of the infection and transmission of cholera.

> *The strength of his model derived from its ability to use observed phenomena on one scale to make predictions about behavior on other scales up and down the chain. ... If cholera were waterborne then the patterns of infection must correlate with the patterns of water distribution in London's neighborhoods. Snow's theory was like a ladder; each individual rung was impressive enough, but the power of it lay in ascending from bottom to top, from the membrane of the small intestine all the way up to the city itself. (Johnson, Ghost Map)*

John Snow's 1849 Theory & 1855 Evidence

1849: Snow developed theory of infection & transmission

- Based on medical knowledge and study of single events – Horsleydown & Albion Terrace

Fully-developed & modern theory of disease

- Infects & reproduces in the small intestine
- Exits from victim, into water supply
- Infects new victims through drinking dirty water

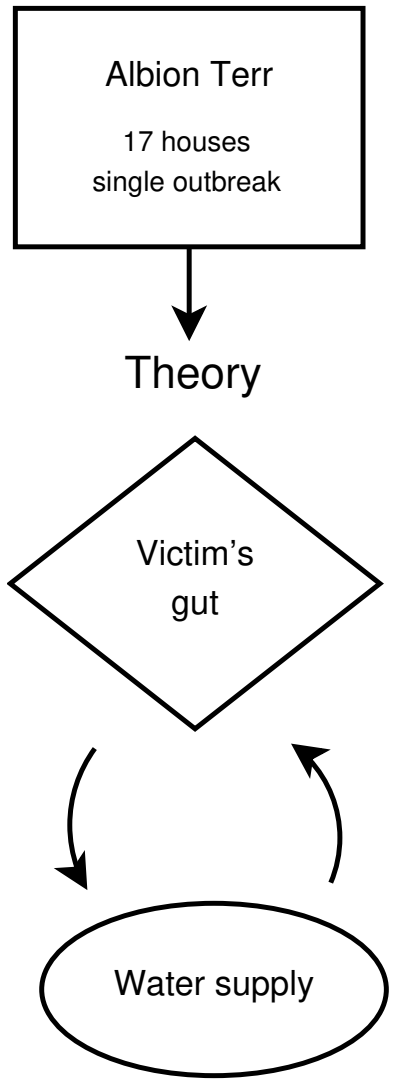
Implications for patterns of infection, across scales

- "from the membrane of the small intestine all the way up to the city itself" (Johnson)

Snow's work grounded by theory

> *Snow had a good idea – a causal theory about how the disease spread – that guided the gathering and assessment of evidence. (Tufte)*

1855: evidence & argument to convince skeptics



Alternative Theories

Miasma (Smells & Airborne)

- Cholera infectious & transmitted through the air
- Generally accepted in mid-1800s

Elevation, Crowding & Class, Others

- Elevation: lower elevation $\rightarrow$ more infection
- Crowding & Class: lower class & crowding $\rightarrow$ more infection

None of these absolutely crazy – correlated with cholera (and dirty water)

- Raw sewage associated with bad smells & dirty drinking water
- Lower class associated with crowding & poor sanitation

Other non-infectious theories (I won't seriously consider)

- Emanations from the ground
- Plague burying-pit near Broad Street pump

Data & Hypothesis Testing

Data or Evidence Blocks

Albion Terr

17 houses
single outbreak

Broad St

~10 sq blocks
2wks, 700 deaths

South London

summer/fall 1854
~400k subjects mixed
treated & untreated

Theory & Hypotheses

water & small intestine | miasma (airborne) | elevation, class, ...

Hypothesis or Testing Blocks

Albion Terr

Narrative

Broad St

Map | Cases | Contin

South London

Diff-in-Diffs | Mixing

No sub-district pop | With sub-district pop

Locations of Events & Data

Data & Hypotheses – Summary

Data	Summary	Statistical Testing	Theory	Refute?
Albion Terrace, 1849	17 houses, narrative	None	Water Miasma	NO YES
Broad St – Hampstead	Single case, "Far from pump but died"	None	Water Miasma	NO YES
Broad St – workhouse	Single cases, "Close to pump but survived"	None	Water Miasma	NO YES
Broad St – 500 residents	Infection rates – 500 residents drink y/n	Contingency Table	Water Miasma	NO YES
S London 1849 vs 1854, 480k people	Mortality rates by supplier	Diff-in-diffs	Water Miasma	NO YES
S London quasi-randomized	Mortality rates supplier	RCT	Water Miasma	NO YES

Data & Hypotheses – Detail

Data	Summary	Statistical Testing	Theory	Refute?	Comment
Albion Terrace, 17 houses & 20-25 deaths, 1849	17 houses infected, surrounding not	None	Water Miasma	NO YES	Sewage leaked into shared water supply after storm. Crucial for *developing* theory
Broad St – Susannah Eley (Hampstead, 1 person)	Single Case, "Far from pump but died"	None	Water Miasma	NO YES	Water bottles shipped to Hampstead by sons
Broad St – St. James workhouse (535 people, 5 deaths)	Counterexample? "Close to pump but survived"	None	Water Miasma	NO YES	In-house well
Broad St 500 residents, categorized by drinking & illness	Infection rates differ by pump drinking	Contingency Table	Water Miasma	NO YES	
S London 480k people, 1849 vs 1854 diff-in-diffs, aggregate sub-district	Mortality rates differ by water supply, not other characteristics	Diff-in-diffs, linear & count regressions, error analysis	Water Miasma	NO YES	Lambeth Water Co changed to clean water 1852, $\Rightarrow$ control / treatment DiD design
S London 480k people, direct District / sub-district comparison, quasi-randomized	Mortality rates differ by water supply company	RCT, Count regressions, detailed error analysis	Water Miasma	NO YES	Mixing of water co customers, control / treatment, effectively randomized

Further Consideration of Data & Hypotheses

Only summarizes Water vs Miasma

- Data also rejected other alternatives: class, crowding, elevation, weather – in fact virtually any we can think of

Must carefully consider *how* data rules out confounding factors

- Broad St residents & contingency: most factors such as crowding & class
- S London DiD: many factors such as weather, elevation, crowding
- S London Quasi-Randomized: most factors such as class, age, income, weather, elevation, crowding, ...

Susannah Eley & Miasma – Data Does *Not* Tell Us The Answer

Data tells us *nothing* – must use judgment, logic, intuition, knowledge of the world

- Cholera Commission rescued miasma with (in our eyes) ridiculous auxiliary hypothesis: "[pump's] impure waters having participated in the atmospheric infection of the district"
- Imre Lakatos & "Core" vs "Protective Belt" of Auxiliary Hypotheses
- We always need these "auxiliary hypotheses" – but need to judge suitability
- Virtually any "core" theory can be protected by the "protective belt" of auxiliary theories

Far from pump but died

	Water Simple	Water Extended	Miasma Simple	Miasma Extended
Core	Drinking	Drinking	Breathing	Breathing
Auxiliary	P[drink~ distance]	People travel to Broad St	P[breath~ distance]	Water infected by air
Implication	deaths~ distance	deaths~ taste for Broad St	deaths~ distance	deaths~ taste for Broad St
Core Refuted?	YES	NO	YES	NO

Snow's Strength: Theory & Evidence

John Snow's 1855 monograph *On the mode of communication of cholera* provides a valuable example and guide for modern-day researchers in the social sciences, a guide for assembling persuasive evidence of a causal effect. The power of Snow's argument derives from employing the following components, although the final was not really available to Snow at the time:

- Well-articulated theory
- Testing predictions against evidence – Consistent consideration and rejection of alternatives
- Multiple tests and source of data
- Careful and honest assessment of the statistical reliability of the evidence