Experimental Design

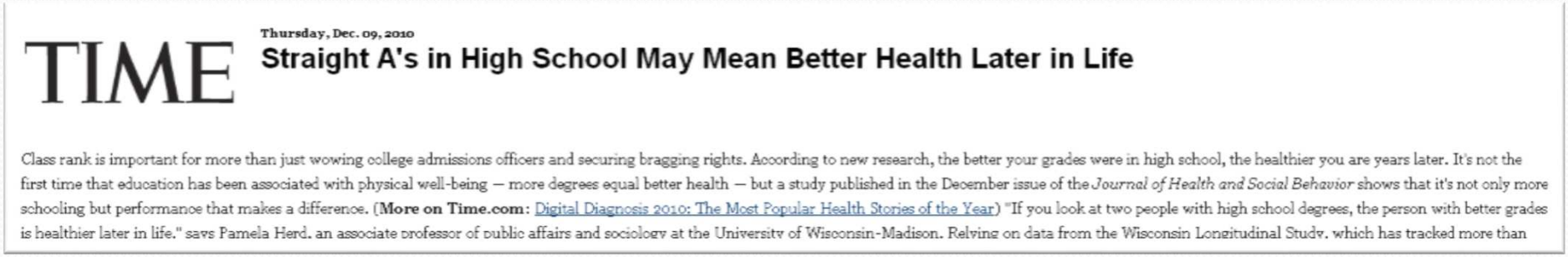
1. Data collection: Conceptual issues
2. Laboratory experiments
3. Natural experiments
4. Establishing causality for invasive species
5. Alternative strategies

Data collection: Concepts

- As we think about the benefits of eradicating or controlling invasive species, we must be cognizant of the difference between *causation* and *correlation*
- Example of *causation*: A 25% reduction East New Britain's cocoa crop caused by the cocoa pod borer will reduce employment in the province by 8200 jobs (Curry et al. 2010)
- Example of *correlation*:

Data collection: Concepts

- As we think about the benefits of eradicating or controlling invasive species, we must be cognizant of the difference between *causation* and *correlation*
- Example of *causation*: A 25% reduction East New Britain's cocoa crop caused by the cocoa pod borer will reduce employment in the province by 8200 jobs (Curry et al. 2010)
- Example of *correlation*:



TIME

Thursday, Dec. 09, 2010

Straight A's in High School May Mean Better Health Later in Life

Class rank is important for more than just wowing college admissions officers and securing bragging rights. According to new research, the better your grades were in high school, the healthier you are years later. It's not the first time that education has been associated with physical well-being — more degrees equal better health — but a study published in the December issue of the *Journal of Health and Social Behavior* shows that it's not only more schooling but performance that makes a difference. (**More on Time.com**: Digital Diagnosis 2010: The Most Popular Health Stories of the Year) "If you look at two people with high school degrees, the person with better grades is healthier later in life," says Pamela Herd, an associate professor of public affairs and sociology at the University of Wisconsin-Madison. Relying on data from the Wisconsin Longitudinal Study, which has tracked more than

Data collection: Concepts

- As we think about the benefits of eradicating or controlling IAS, we must be cognizant of the difference between *causation* and *correlation*
- Example of *causation*: A 25% reduction East New Britain's cocoa crop caused by the cocoa pod borer will reduce employment in the province by 8200 jobs (Curry et al. 2010)
- Example of *correlation*:

Data collection: Concepts

- As we think about the benefits of eradicating or controlling invasive species, we must be cognizant of the difference between *causation* and *correlation*
- Example of *causation*: A 25% reduction East New Britain's cocoa crop caused by the cocoa pod borer will reduce employment in the province by 8200 jobs (Curry et al. 2010)
- Example of *correlation*:

ScienceDaily
Your source for the latest research news

Web address:
http://www.sciencedaily.com/releases/2008/09/080922090801.htm

Print this pag

Sexism Pays: Men Who Hold Traditional Views Of Women Earn More Than Men Who Don't, Study Shows

ScienceDaily (Sep. 22, 2008) — When it comes to sex roles in society, what you think may affect what you earn. A new study has found that men who believe in traditional roles for women earn more money than men who don't, and women with more egalitarian views don't make much more than women with a more traditional outlook.

Timothy Judge, PhD, and Beth Livingston from the University of Florida, analyzed data from a nationally representative study of men and women who were interviewed four times between 1979 and 2005. A total of 12,686 people, ages 14 to 22 at the beginning of th study, participated; there was a 60 percent retention rate over the course of the study.

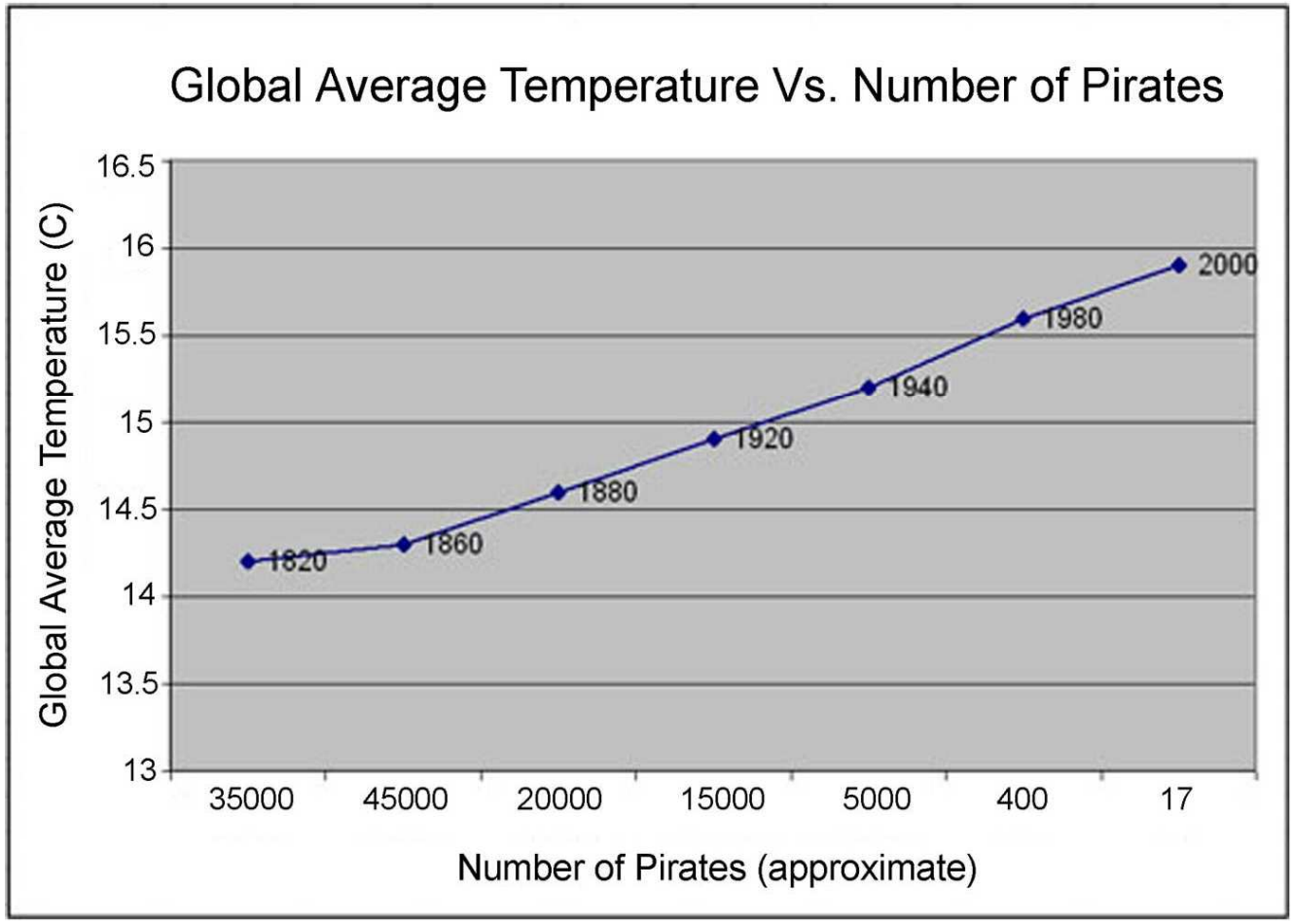
Data collection: Concepts

- As we think about the benefits of eradicating or controlling invasive species, we must be cognizant of the difference between *causation* and *correlation*
- Example of *causation*: A 25% reduction East New Britain's cocoa crop caused by the cocoa pod borer will reduce employment in the province by 8200 jobs (Curry et al. 2010)
- Example of *correlation*:

Expert opinion: Take care as you drive home from work because most traffic accidents happen within 10 km of your house.

Data collection: Concepts

STOP GLOBAL WARMING: BECOME A PIRATE



Data collection: Concepts

- Average per capita income from cocoa farming in East New Britain (where cocoa pod borer is present)
  = 714 kina
- Average per capita income from cocoa farming in Ghana (where cocoa pod borer is not present)
  = 300 cedi (which is 339 kina)
- Therefore, eradicating cocoa pod borer will reduce average incomes in East New Britain by 375 kina

Establishing Causality: Laboratory Experiments

- In **randomized trials** or **laboratory experiments**, we select a sample representative of the population and randomly assign people to **treatment** and **control** groups
  - Require villagers to draw random numbers;
    Require odd-numbered children to eat fish (treatment) and prevent even-numbered children from doing so (control)
  - Require odd-numbered children attend religious services for their whole lives (treatment) and prevent even-numbered children from doing so (control)
  - Infest odd-numbered people's cocoa crops with CPB (treatment) and prevent CPB infestation in even-numbered people's crops (control)
- We can assess the effect of treatment on some outcome by a simple comparison of mean outcomes between the treatment and control groups
- Challenges: practicality, feasibility, cost, ethics

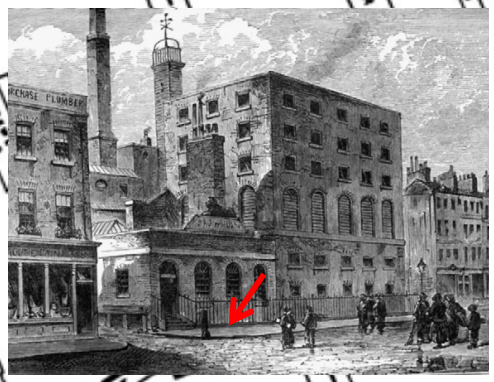
Establishing Causality: Natural Experiments

- In some cases, nature provides laboratory conditions
- Case study: Nineteenth-century London
  Cholera killed thousands in waves of epidemics
  - There was much speculation regarding the cause
  - The leading theory was that “miasma” was the primary source of cholera and that carrying herbs on one’s body provided immunity

Establishing Causality: Natural Experiments

- Dr. John Snow noticed some things about cholera:
  - Cholera incidence was higher among the poor, who lived in crowded and often unhygienic conditions
  - Some cholera outbreaks were extremely localized, affecting people living in one building, but not the next
  - Primary symptoms of cholera are vomiting and diarrhea
- With this evidence, Snow speculated that cholera was spread through an organism that passed through body and into the water supply

Establishing Causality: Natural Experiments



- During the epidemic of 1853-1854, he examined a spatial cluster of cholera cases

Establishing Causality: Natural Experiments

- Snow observed that a number of water companies took their water from the Thames, which was highly contaminated with sewage
- There were three major water companies in London:
  - Southwark and Vauxhall drew their water from below the main sewage discharge for London
  - Lambeth withdrew its water from above the discharge
- Suppose that Snow found that clients of Lambeth were statistically less likely to get cholera than clients of Southward and Vauxhall
  - Would this convince you that cholera was a waterborne disease spread through sewage-contamination?

Establishing Causality: Natural Experiments

- It turns out that Londoners were subscribed to water companies in a close to a random manner. Rich and poor were equally likely to take their water from each of the companies
- The companies each served all of London; they didn't divide the city up into sections with exclusive service by one company to one section of the city

"The mixing of the supply is of the most intimate kind. The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies."

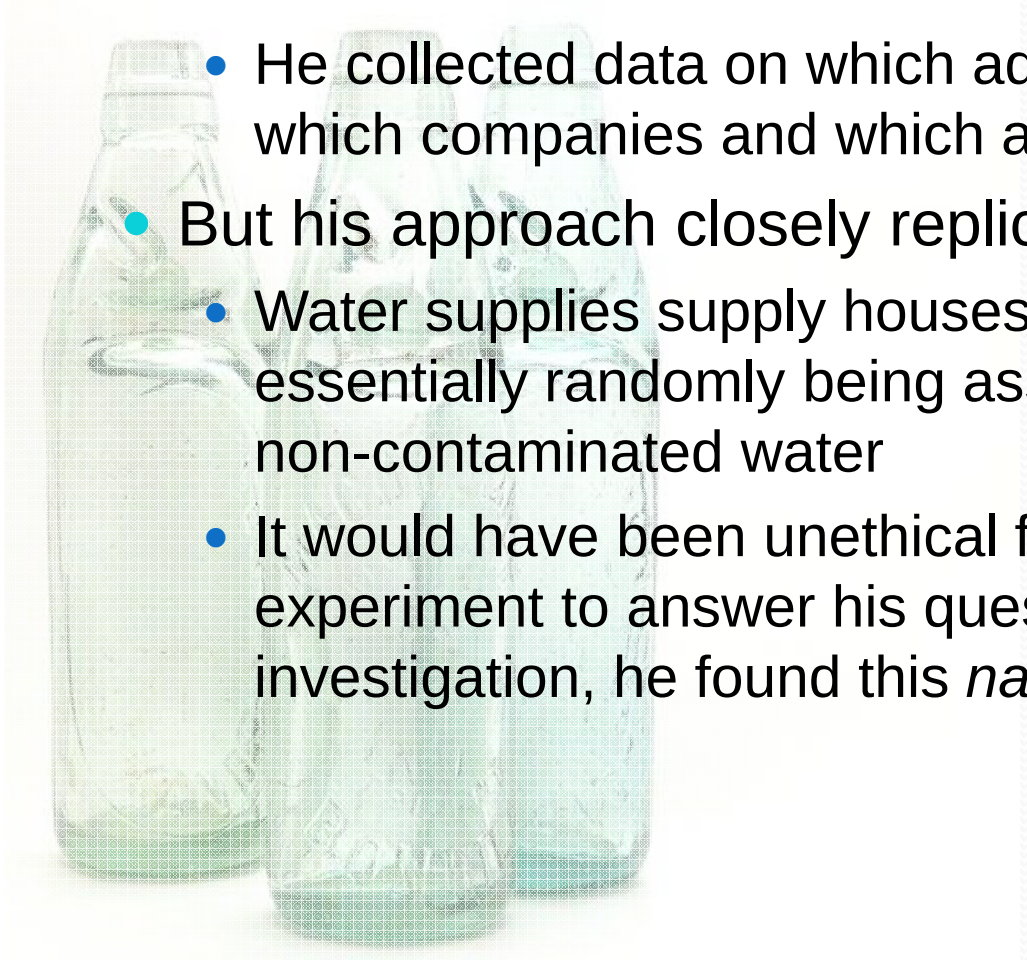
Establishing Causality: Natural Experiments

"The experiment, too, was on the grandest scale. No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into two groups without their choice, and in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and amongst it, whatever might have come from the cholera patients, the other group having water quite free from such impurity."

- Snow refers to this as "the experiment"
- With a true laboratory, one would randomly assign odd-numbered people to drink sewage-contaminated water (treatment) and even-numbered people to drink non-contaminated water (control)
- Then, one would simply compare mean cholera rates between the two groups

Establishing Causality: Natural Experiments

- Snow's data are *observational*
  - He collected data on which addresses got their water from which companies and which addresses had cholera deaths
- But his approach closely replicates a *laboratory experiment*
  - Water supplies supply houses in a near random way, so it is essentially randomly being assigned to drink contaminated vs. non-contaminated water
  - It would have been unethical for Snow to administer a literal experiment to answer his question, but with some careful investigation, he found this *natural experiment*



Establishing Causality: Natural Experiments

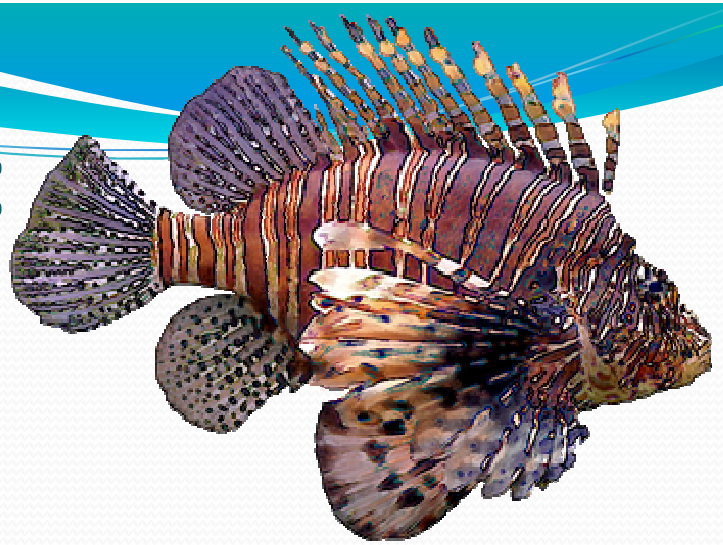
- Table IX from Snow's Results

Company	# Houses	# Cholera Deaths	# Cholera Deaths / 10,000 Houses
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Establishing Causality: Lessons courtesy of John Snow

- Lesson 1: Good empirical science can be done without high-tech techniques
- Lesson 2: The truth is out there, but unless we have a laboratory, we need to be clever to uncover exogenous variation in the variable of interest
  - This is true whether the problem is dirty water or IAS!

Establishing Causality: Invasive Species

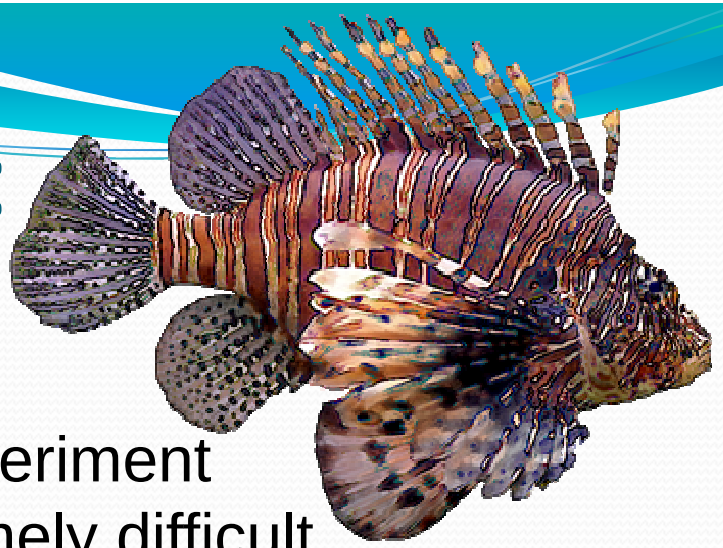


- IAS *may* themselves be a natural experiment
  - e.g., suppose that IAS reach a community *after* you have already collected baseline data
    - The community is both treatment (after IAS) and control (before IAS)
    - Challenge: Did anything else in that community change over time?
  - e.g., suppose a natural barrier prevents an IAS from reaching one village from another in the same community
    - The village with IAS is treatment and the other is control
    - Challenge: Are the villages identical in every way apart from the presence of invasive species?

Establishing Causality: Invasive Species

- However, if IAS are not randomly distributed, then we *cannot* establish causality
  - Suppose that IAS only travel 5km from the coast
    - Are interior communities the same as coastal communities?
  - Or that they don't climb above a certain elevation
    - Are mountain communities the same as lowland communities?
  - Or that they only affect banana trees
    - Are banana growers the same as other farmers?
- We also cannot establish causality if communities change behaviour in anticipation of invasive species
  - e.g., suppose that loggers overharvest because a plant pathogen is expected to affect the area

Establishing Causality: Invasive Species



- If you do not have a laboratory experiment or a natural experiment, it is extremely difficult to establish causation
- Without causation, we generally talk about **correlation**:
  - Piracy is negatively correlated with average global temperatures
- Without causation, we cannot assign precise values:
  - Communities with iguanas may have 10% lower income from tourism, but that may be true of those places even in the absence of iguanas
- But correlation is still helpful! So don't give up if you cannot identify a natural experiment

Alternative Strategies

1. Laboratory experiment
2. Natural experiment
3. Retrospective data from administrative records
4. Econometric methods

   e.g., IV estimation, propensity score matching, regression discontinuity (requires considerable data and statistical knowledge)

5a. Secondary data sources

5b. Primary data from surveys