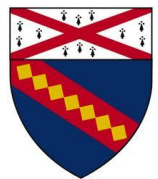


A defense of ecological analysis

Dan Weinberger, PhD
Associate Professor of Epidemiology



Yale SCHOOL OF PUBLIC HEALTH

What is an ecological analysis?

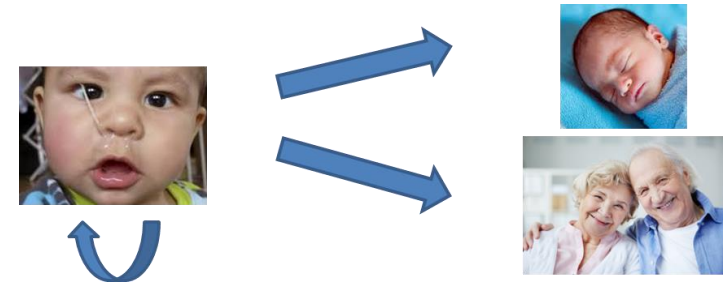
- Correlation of aggregate data
 - Time trend analyses (how do disease rates change after introduction of a vaccine?)
 - Spatial analyses (association of ZIP code-level incidence of lung cancer with ZIP code level smoking rates)



Why is ecological analysis essential?

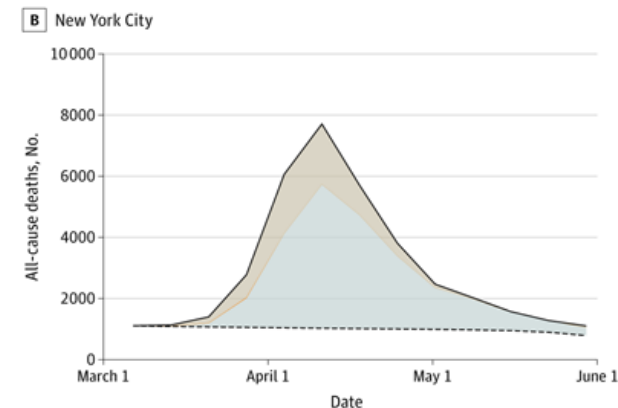
- Some phenomena act at the population level or can only be measured at the population level, so individual-level analysis cannot capture it:

Vaccines provide direct protection to individual and disrupt transmission, providing an additional community benefit



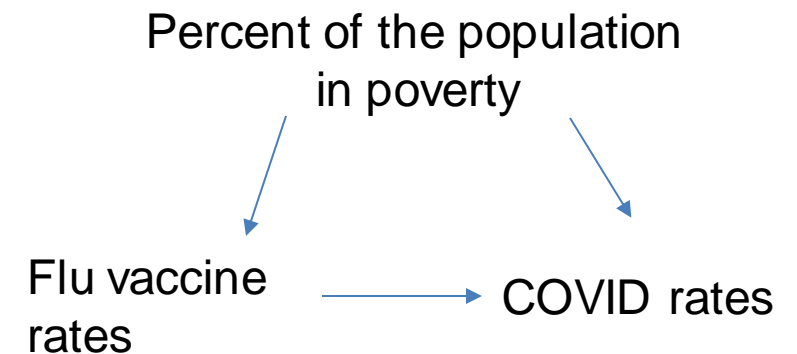
COVID-19 Pandemic had both individual-level and societal impact that affected health

- Infection
- Overwhelmed healthcare
- Other social/environmental determinants



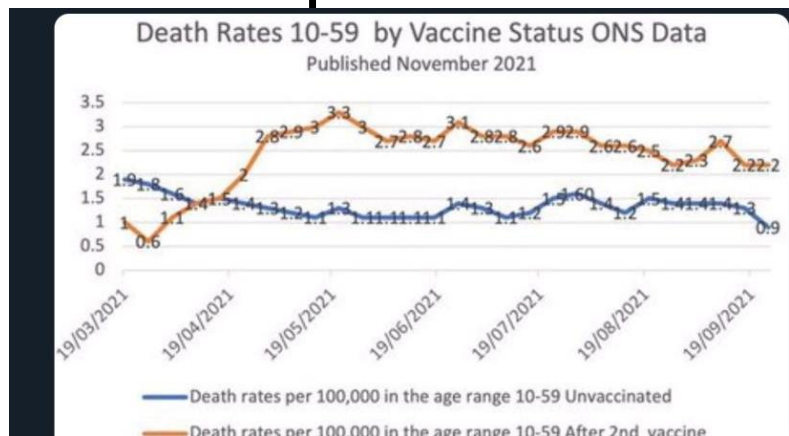
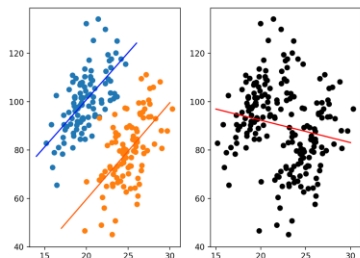
Why do people hate ecological analyses?

- Often not done in a rigorous way
 - Insufficient adjustment for confounding (or inability to adjust for relevant confounders);
 - Association of hot chocolate sales with influenza rates



Why do people hate ecological analyses?

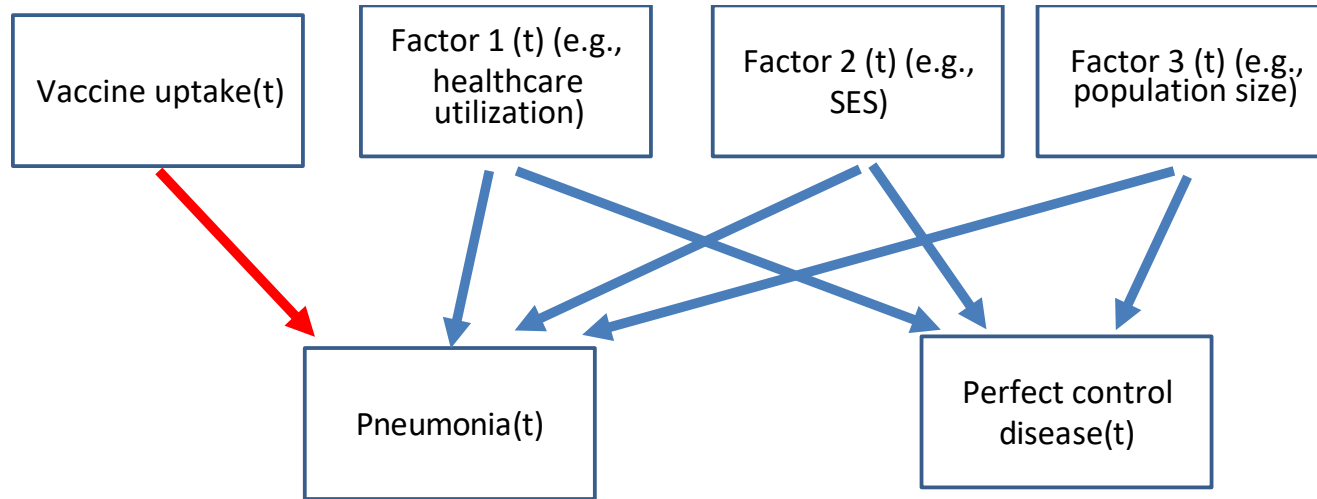
- “The ecological fallacy”: inferring individual level effects from aggregate associations
 - Provinces with higher % Protestant had higher suicide rates in Europe in 19th century than predominantly Catholic provinces \neq Protestants more likely to commit suicide
- Simpson’s Paradox: relationships at individual level flip when viewed in aggregate



Age Band	Unvaccinated			Vaccinated			Unvaccinated / Vaccinated
	People	Deaths	Deaths per 100k	People	Deaths	Deaths per 100k	
10-29	5,000,000	40	0.8	5,000,000	20	0.4	200%
30-59	1,000,000	60	6.0	10,000,000	300	3.0	200%
10-59	6,000,000	100	1.7	15,000,000	320	2.1	78%

*Illustrative numbers to show Simpson's Paradox at work.
 Within both age groups the unvaccinated death rate is twice the vaccinated death rate.
 However, in the combined age group the unvaccinated death rate is lower.
 This is because the average unvaccinated person is much younger than the average vaccinated person.*

Causation from ecological data



Regression: $E(\text{pneumonia cases}_t) = b_0 + b_1 * \text{Perfect_control}_t$

Identify confounders at the **same level of measurement**

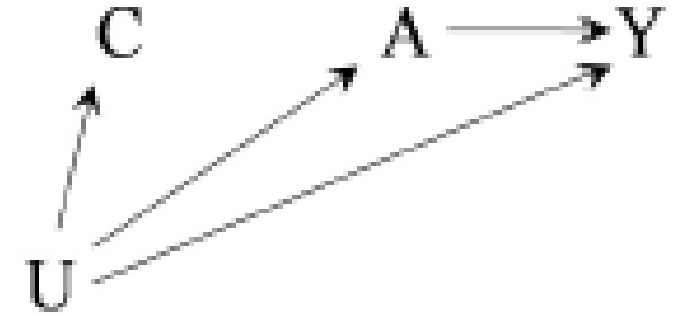
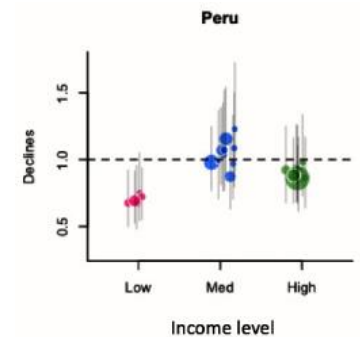
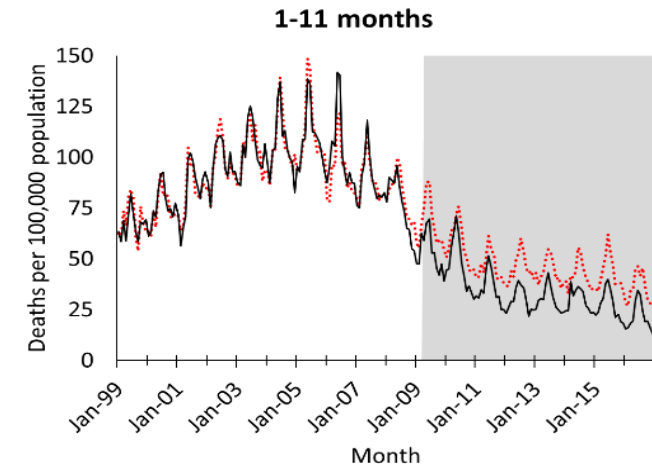


Figure 7.13

Difference-in-difference or synthetic controls approaches

General approaches to improve the rigor of ecological studies

- Control variables
 - E.g., ‘synthetic controls’ approach: use other causes of death to adjust for trends unrelated to a vaccine
- Negative control outcomes
 - outcomes that shouldn’t be affected by the change
 - Evaluate changes that happen just *before* intervention
- Incorporate time series data from multiple spatial units
- Combine ecological + individual level analyses



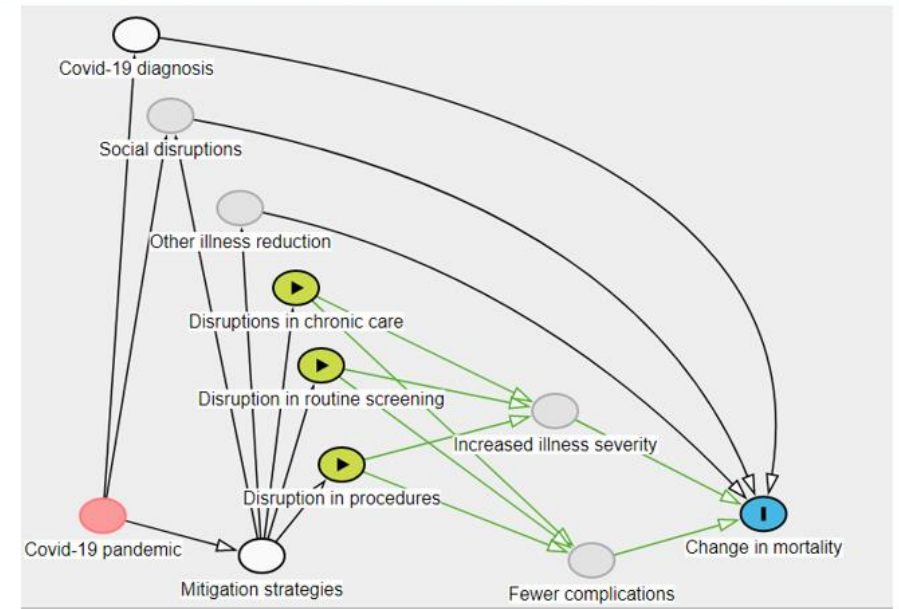
The flip side.

- “individualistic fallacy” in which the major population determinants of health are ignored and undue attention is focused on individual characteristics.

Combining ecologic and individual-level analyses in DCNP

- There are many ecologic variables of relevance for DCNP (measures of hospital capacity and disruption, community transmission rates)
- Also relevant individual-level variables (e.g., loss of control of chronic conditions)

DCNP Conceptual Framework



Courtesy Caroline Korves

Example: Pandemic-related increases were more pronounced for women in the VA (now: why?)

