Stat547L: Spatio–temporal methods in environmental epidemiology

Lecture 21

SPECIAL TOPICS:

Lecture 21. Environmental health risk assessment

Lecture 22. Measurement error

Lecture 23. Designing good monitoring networks

Lecture 24. Dealing with high dimensional data

Outline: Lecture 22

1. **Environmental health risks**
2. **Uncertainties on the road to causality**
3. **Roadblocks!**
4. **Circumenting them**
5. **Circumenting them**

Environmental health risk

This lecture presents an overview of the obstacles analysts face the assessment of environmental risk assessment. In particular, it focuses on the problem of establishing causal links between environmental hazards and adverse health outcomes such as morbidity, mortality, school absences due to sickness and hospital admissions.

Respiratory health

Health effects of ambient air pollution:

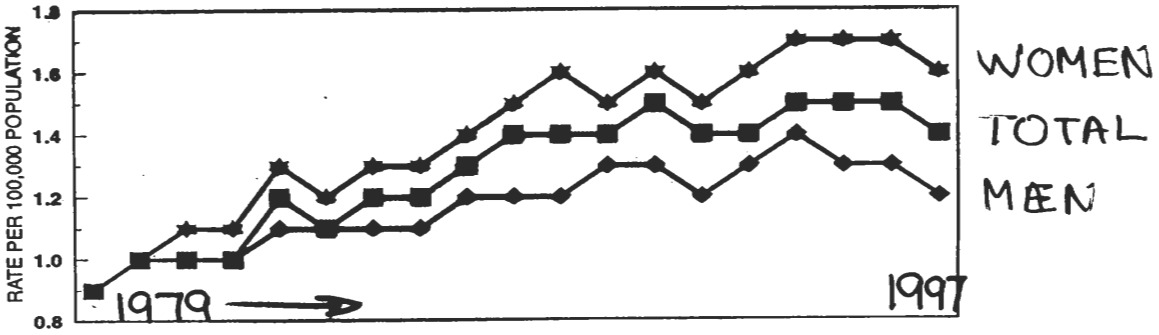
- **Fairly certain:**
  - chronic bronchitis - caused and exacerbated
  - asthma - exacerbated
  - emphysema - exacerbated
  - pulmonary function - reduced
  - cough - caused
  - death - contributor
- **Possible:**
  - emphysema - caused
  - infection - weakened defenses
  - lung disease - occupational settings

Trends

Lots of available evidence points to a determination in health due to exposure to environmental hazards. Example:

ASTHMA DEATH RATES ON THE RISE!

FIGURE 1: ASTHMA: AGE-ADJUSTED DEATH RATE BY SEX, 1979-1997



YEAR	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
BOTH SEXES	0.9	1.0	1.0	1.0	1.2	1.1	1.2	1.2	1.3	1.4	1.4	1.4	1.5	1.4	1.4	1.5	1.5	1.5	1.4
MALE	0.9	1.0	1.0	1.0	1.1	1.1	1.1	1.1	1.2	1.2	1.2	1.3	1.3	1.2	1.3	1.4	1.3	1.3	1.2
FEMALE	0.9	1.0	1.1	1.1	1.3	1.2	1.3	1.3	1.4	1.5	1.6	1.5	1.6	1.5	1.6	1.7	1.7	1.7	1.6

SOURCE: NATIONAL CENTER FOR HEALTH STATISTICS: ANNUAL SUMMARY OF VITAL STATISTICS, 1979-1997

Particulate air pollution

Particulates seem particularly nasty. They may consist of just “road dust” (including pollen + other allergens). More lethal brews come from combustion.

PM sources

Primary sources:

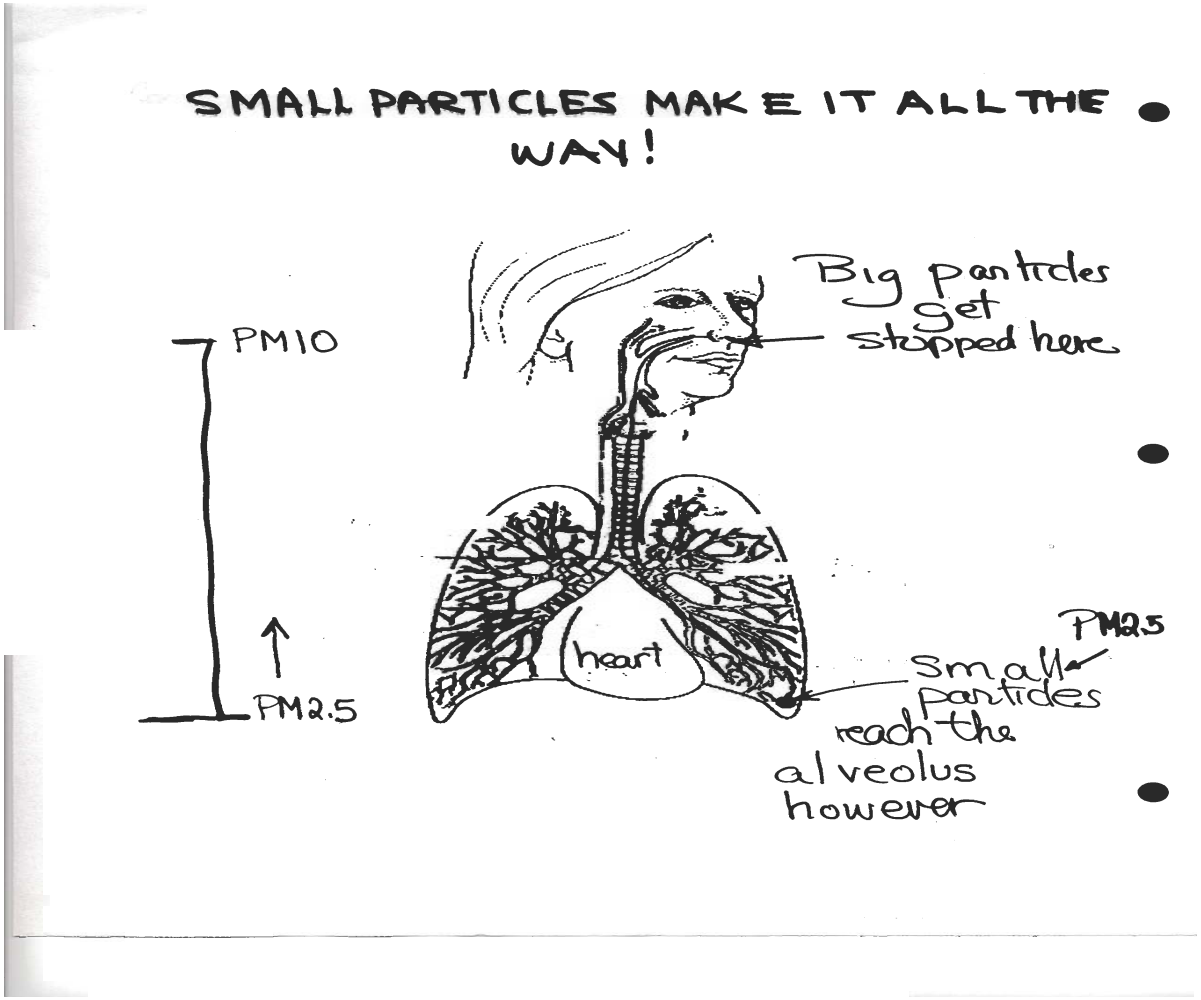
- motor vehicles
- power plants
- wood fires
- industrial processes
- tobacco smoke
- fireplaces
- gas cooking stoves

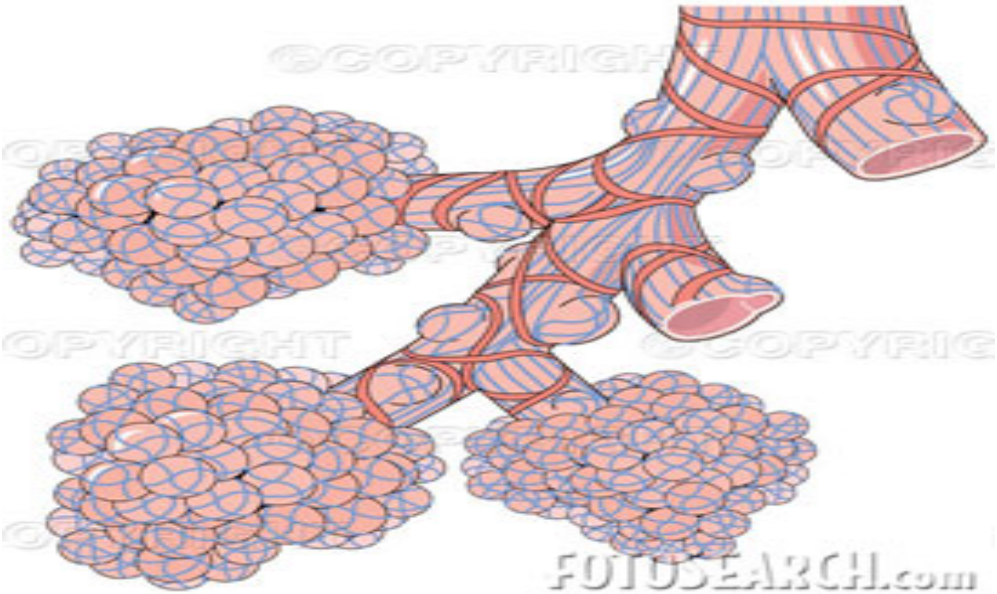
Secondary sources:

- photochemical transformations of SO_2, NO_x & volatile organic compounds from fuel burning

- PM_{10} are inhalable particles, size 10 microns or less, $PM_{2.5}$ to those under 2.5.

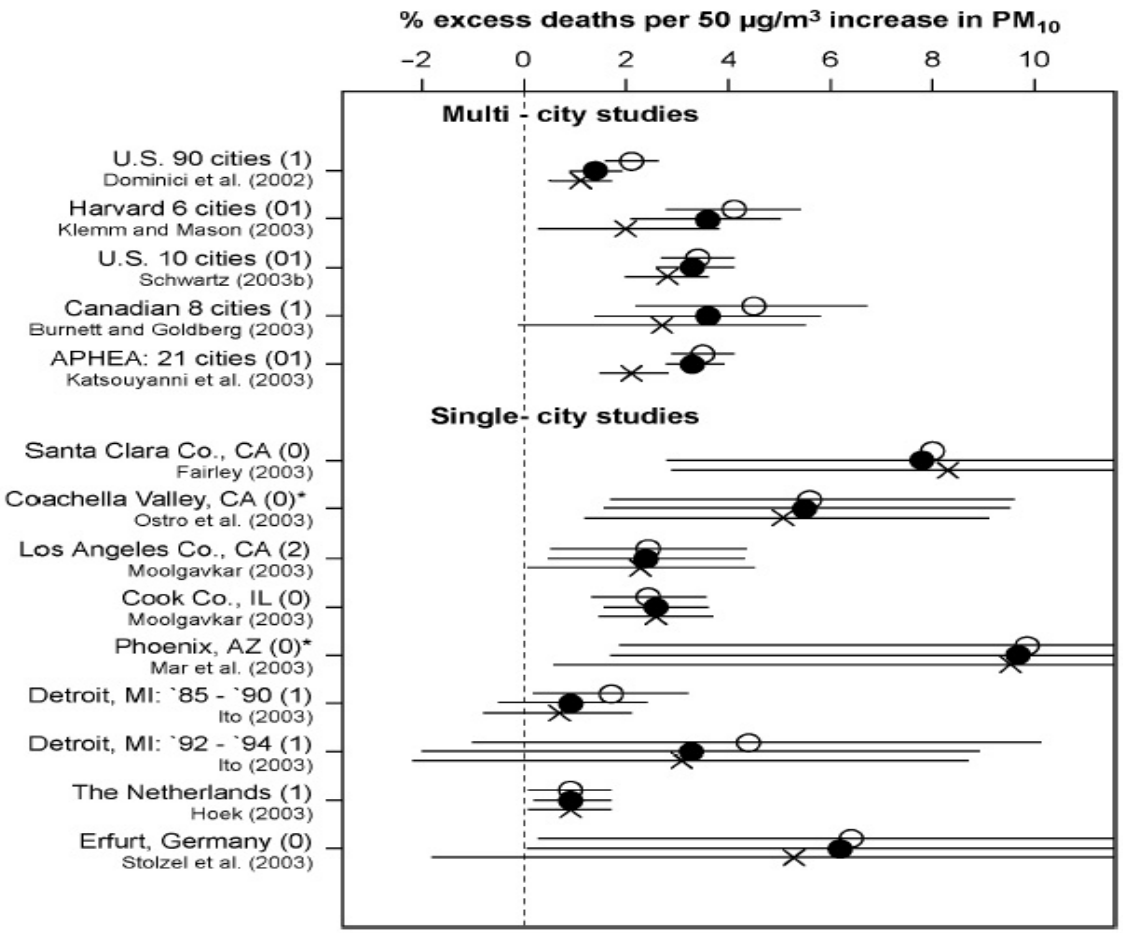
Particulates & lungs




FOTOSEARCH.com

- Measured in counts (ppm) or mass ($\mu g\ m^{-3}$). Counts can be large but mass small when particles are fine. Dangerous! Can get all the way down into the lung. 10 or 50 ($\mu g\ m^{-3}$) is one natural scale on which the relative risk of PM_{10} particulates is considered.
- Consistent, substantial body of evidence implicates haze, smoke, soot, dust in health risk analysis. Many spatial epidemiological studies since the mid 1990s. They show a consistent pattern of association between particulates and morbidity as well as mortality.

PM's relative health risk studies and apparent causation



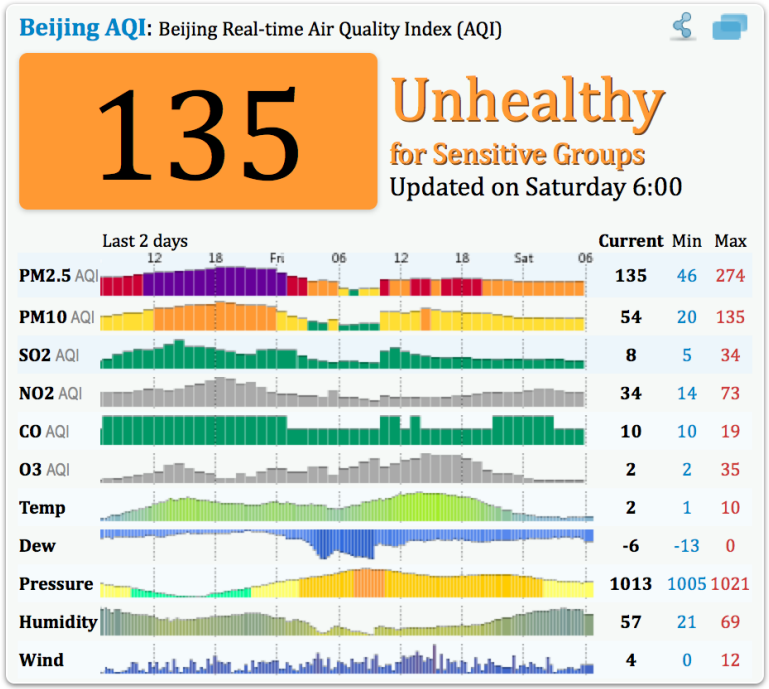
USA's $PM_{2.5}$ *standards*

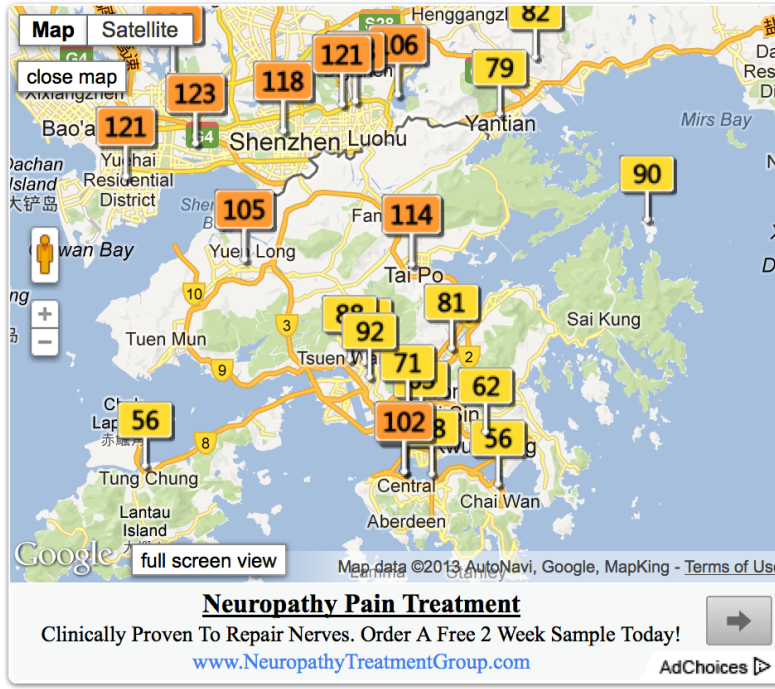
Apparent causation has led to imposition of National Ambient Air Quality Standards (NAAQS) in the US under the Clean Air Act (1970) based on averages over "designated" monitoring sites:

Annual $PM_{2.5}$ standard: 3 year average of the annual mean concentrations for designated monitoring sites in an area must be ≤ 15.0 (μm^{-3}).

24 hour standard: 3 year average of annual 98th percentile of values at designated monitoring sites in an area must be ≤ 35 (μm^{-3}).

In Beijing: Mar 22, 2013





Very Unhealthy	**201 to 300**	**Health warnings of emergency conditions. The entire population is more likely to be affected.**

A case for Causality?

Uncertainties:

- above all no smoking gun!
- epidemiological studies observational, not controlled, randomized. Cannot prove causation
- studies are commonly based on aggregate data (**ecological effect**)
- commonly involve large **errors of measurement**

Components of measurement error

- use of aggregate not individual data
- use of ambient, not individual exposure data (what about dose?)
- difference between true and measured ambient level

NOTE: Effects (relative risks) are small and subtle - hard to separate from other effects - so measurement error can be serious.

A potential smoking gun!

Hypothesis:

Inhaled PM (airborne particulates) initiate production by lung cells of inflammatory mediators and oxidants that become bloodborne and target the cardiovascular system. (MESA Study at U of Washington.)

Going deeper

An experimenter finds a strong relationship between daily log (*ozone*) concentrations and daily log (*Hospital Admissions*) for a number of central of census subdivisions.

CONCLUSION: High ozone concentration levels cause people to go to the hospital?

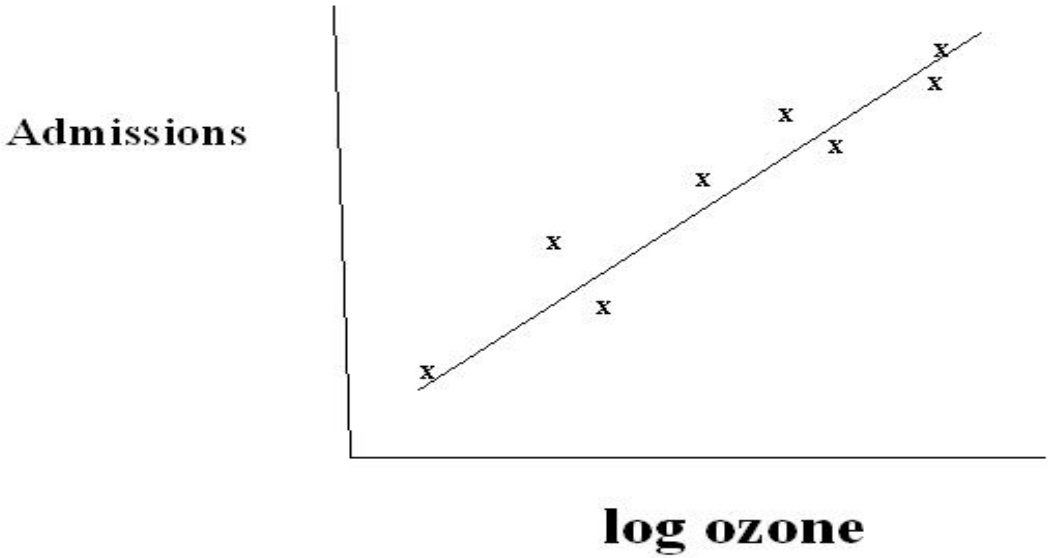
QUESTION: Why is this conclusion unjustified?

ANSWER:

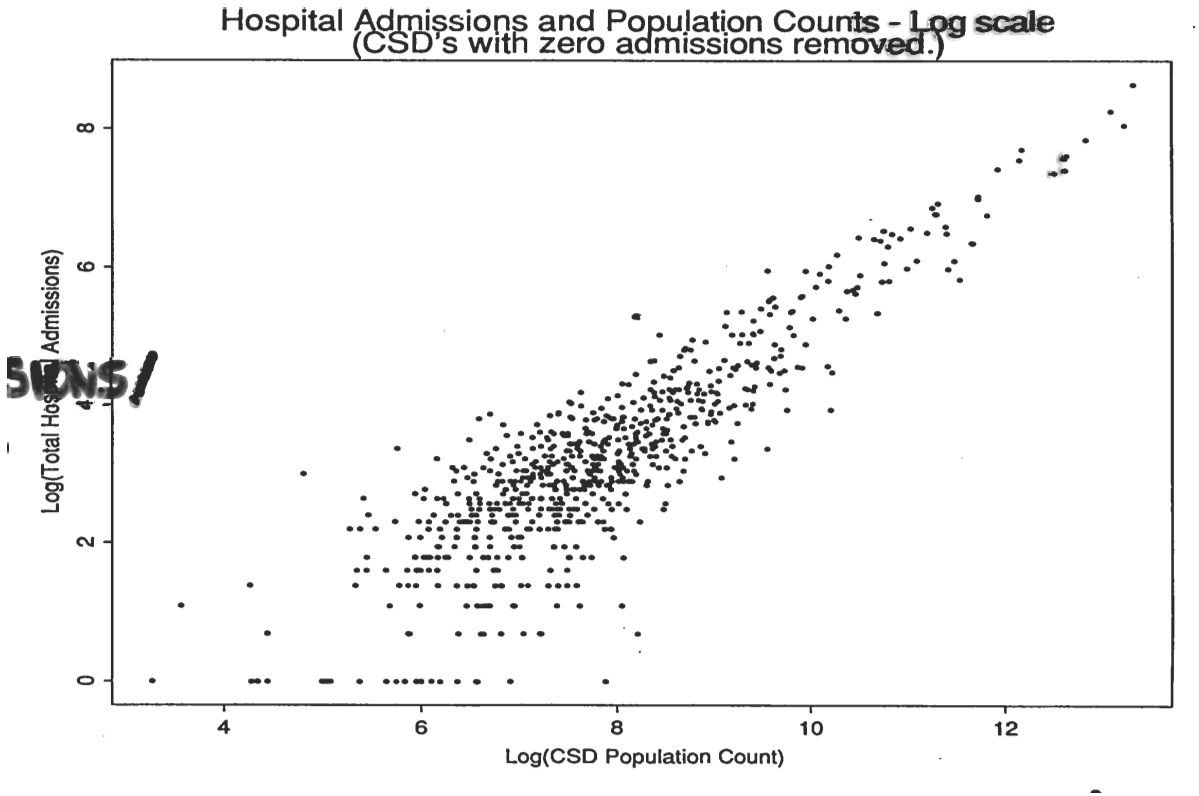
- Ecological fallacy: conclusion based on aggregate, not individual data
- Observational studies always subject to possible confounding. (Here **population size** an obvious confounder.)

So measurement error is not the only problem.

Hypothetical Example



Hospital admissions are strongly influenced by population size. The following figures show relationship for census subdivisions.



Going deeper still: What is Causality?

Subject to much debate. One possible definition is by the:

> **COUNTERFACTUAL ARGUMENT:** A causes B if B changes when A changes and everything else is held constant.

PROBLEM: Changing particulate levels would need a lot of other changes!

So instead, try an operational definition of causality

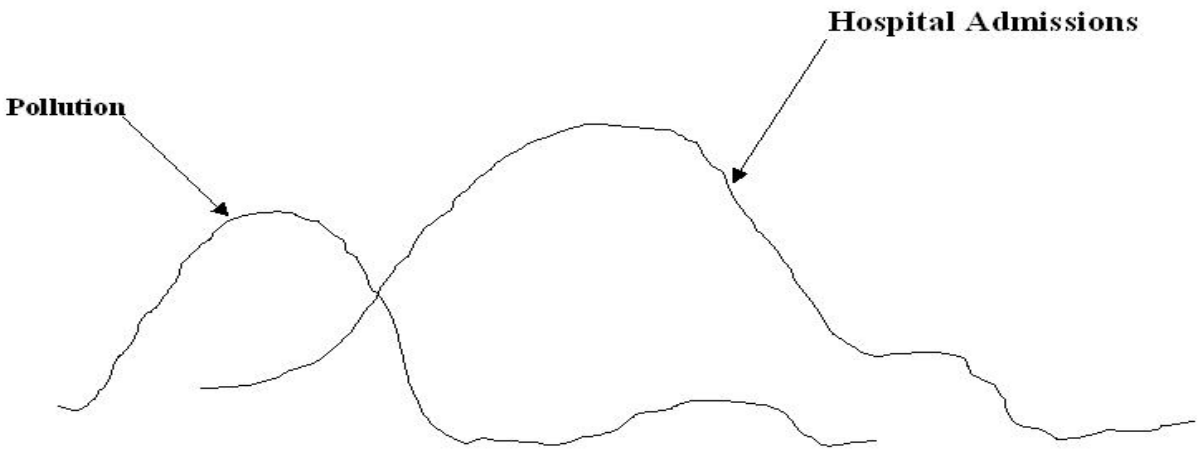
BRADFORD HILL CRITERIA:

The following conditions serve as a surrogate for definition of causality:

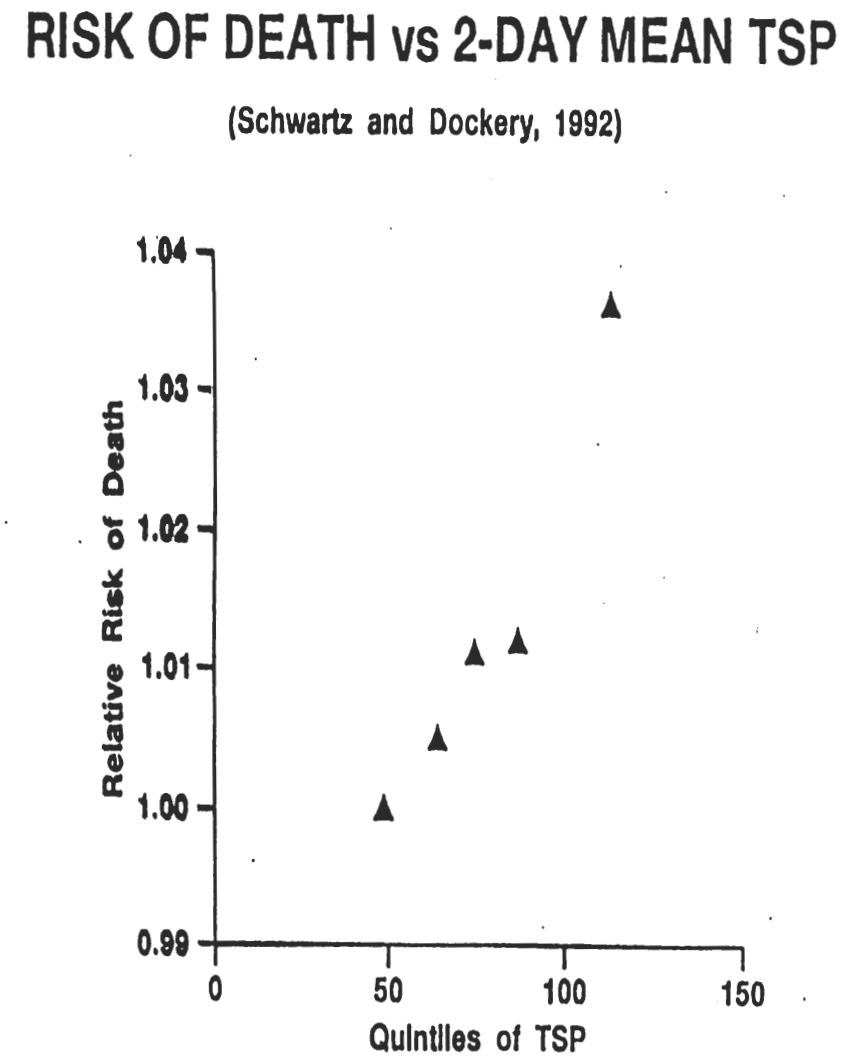
1. Strength of association – if A happens then B usually happens (not enough on its own)
2. Specificity – if A does not happen, B usually does not either (not enough on its own)
3. Biological plausibility

More on the following slides.

4. Temporal gradient



5. Biological gradient



6. Consistency of results

Relative risk of respiratory health outcome vs particulate increase (Goldberg 1996, J Aerosol Medicine)

Place	**Particulate**	**% RR increase**
New York (63-72)	SS	1.5%
Cincinnati (77-82)	TSP	1.2%
Philadelphia (73-80)	TSP	1.1%
Birmingham (85-88)	PM_{10}	1.1%
Utah Valley (85-89	PM_{10}	1.4%
Cracow (77-89))	PM_{20}	0.2%
Athens (84-89)	BS	0.8%
Beijing (89)	TSP	1.1%

Taking the statistical road to causality

Completely randomized experiment. E.g. assign 1/2 of the subjects at random to ozone gas chamber and 1/2 to placebo chamber.

PROBLEMS:

- Ethics!
- With small # of subjects “confounders” could align themselves with the treatment. Completely randomized block designs could mitigate but not avoid this problem
- To control for confounders will mean experimenting in unrealistic conditions. E.g. strawberries grow well in an ozone tent on UBC campus. That would not mean high ozone levels predict good crops in the Fraser Valley!

Realistic alternative – design observational studies using randomized design principles.

- select “cases” from regions with high pollution and “controls” from regions with low pollution.
- use “case-crossover” designs where subjects serve as their own controls (compare affected group on high ozone days with that same group on low ozone days before
- use longitudinal data analysis/designs

But avoid at all costs the:

PHYSICIAN'S FALLACY: assess exposure of those admitted with say asthma attacks to see if they suggested high exposure in the days previous to that on which they were admitted!

Roadblocks on the statistical road: Confounders

Confounders. These can be unmeasured environmental hazards correlated with that of interest. Or other factors such as meteorological. E.g. mortality goes up in hot weather so that it is now regarded as a hazard in a world of changing climate.

Criteria for confounding factor [Rothman and Greenland (1998)]:

1. must be a risk factor for the response.
2. must be associated with the exposure under study in the source population.
3. must not be affected by the exposure or the response. In particular it cannot be an intermediate step in the causal path between the exposure and the response.

Roadblocks on the road: Aggregated data

Ecological fallacy represents divergence of effects at the aggregate level from those at the unit level. Problems are well illustrated in the following fictitious example:

SIMPSON'S PARADOX

	Pollution level	
Deaths (1000s)	**Lo**	**Hi**
Hi	210	508
Lo	150	502

INITIAL CONCLUSION: Pollution is good for you!

Disaggregated data for Regions. Death counts are in 1000s.

	Region 1		Region 2		Totals	
	Pollution level		Pollution level		Pollution level	
Deaths	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**
Hi	200	8	10	500	210	508
Lo	100	2	50	500	150	502

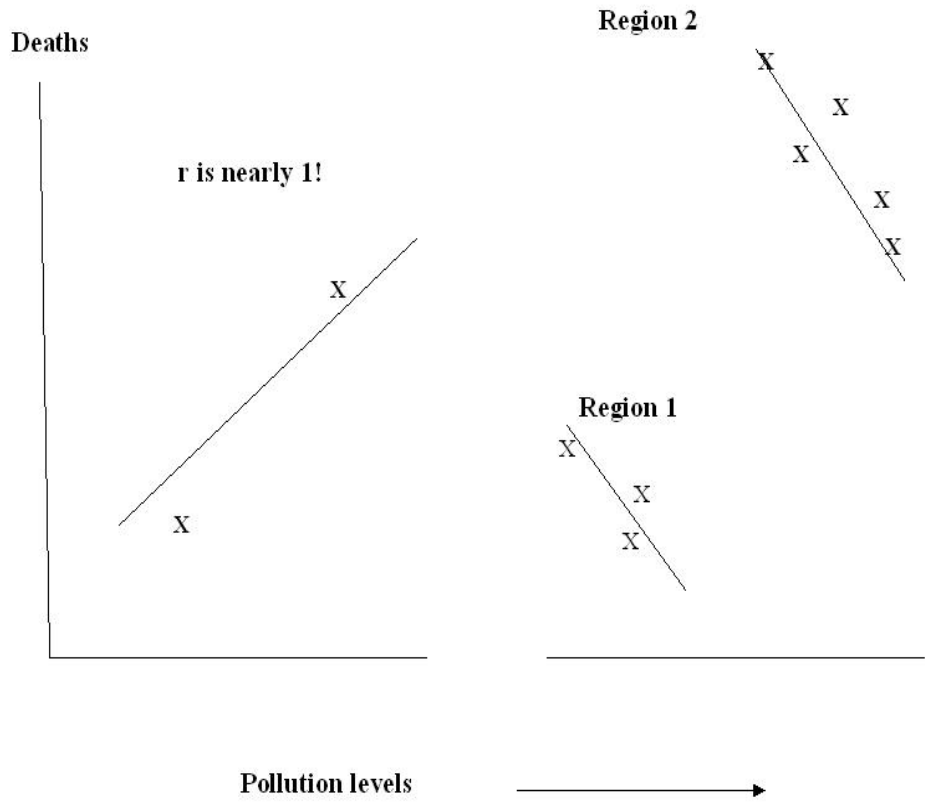
CONCLUSION: Pollution is bad for you in Regions 1 & 2, yet good for you overall!

Confounding and the ecological fallacy

The ecological effect has two potential effects. One is the avoid a fallacious conclusions due to confounding as seen above. The other is unreasonably high levels of association caused by smoothing out the seen at the unit level, the “noise” seen in scatterplots for the individuals.

Graphical demonstration.

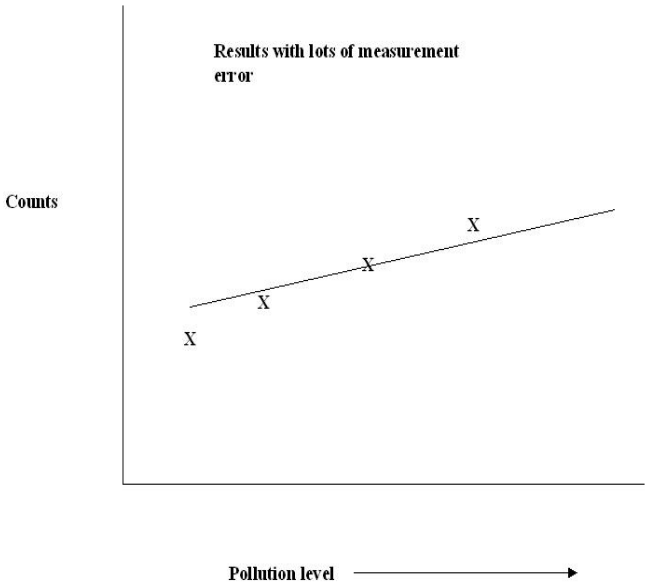
Left panel shows pollution not only that pollution is good for you but that the association is strong ($r = 1$)! Completely misleading results with overconfidence in them. But the right hand panel shows exactly the opposite - with less conviction.

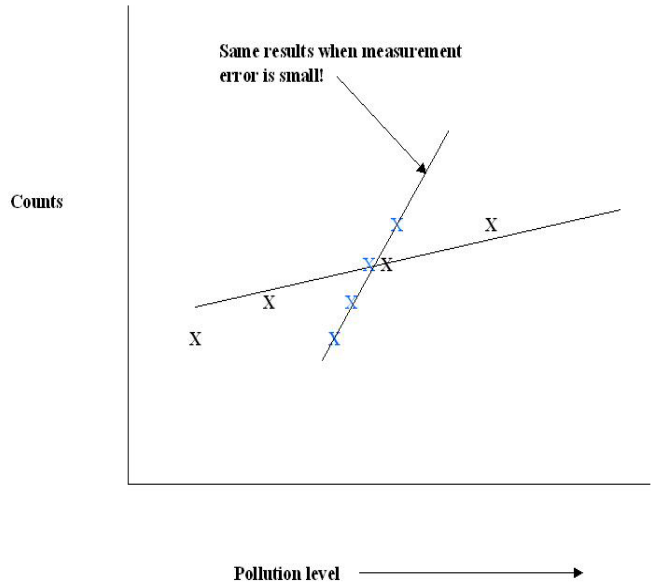


Roadblocks on the road: Measurement Error

Measurement error can have serious and unpredictable effects. (More in Lecture 22.) Generally, its thought to shrink the relative risk toward the null hypothesis of no association. But for nonlinear models, it can go either way!

Graphical demonstration. Here we see in left panel X^* values, which are X values measured with error, plotted on horizontal axis. In the right panel we see X values plotted as well with smaller spread with the same rise in Y. Yields a much larger estimate of β.





Roadblocks on the road: Collinearity

A major issue in environmental epidemiology when seeking cause. The risk factors align themselves.

- Possible that every one of a suite of pollutants can be found to be individually significant, yet none significant when all fitted together.
- Difficult to unravel: temperature and ozone are highly correlated and both are risk factors for mortality. How much should ozone be blamed as a causative agent? **In other words, what is the true cause of death?**

Example: [1]: Daily hospital admissions for respiratory morbidity in census subdivisions for southern Ontario vs O_3 and SO_4. They competed as predictors depending on lags. Next table reveals the competition in stepwise fitting. "L" means days lagged.

	Ranking					
Step	**1**	**2**	**3**	**4**	**5**	**6**
1	O_3 *L2*	SO_4 *L1*	O_3 *L1*	O_3 *L3*	SO_4 *L2*	SO_4 *L3*
2	SO_4 *L1*	O_3 *L1*	O_3 *L3*	SO_4 *L0*	SO_4 *L3*	
3	O_3 *L3*	SO_4 *L3*	SO_4 *L0*			
4	SO_4 *L0*	*Temp L4*				
5	O_3 *L0*					

[1] Zidek et al. [1998]

Avoiding the roadblocks

Contending with confounders.

- Cross sectional studies look at different samples of primary units, eg census subdivisions at successive times.
- Longitudinal studies will keep the same sample and make repeated measures.
- Latter measures long term trends better. Also allows unit changes in health outcomes to be compared directly against its pollution changes reducing possible intervention of confounders!!

Coping with measurement error This arises in environmental epidemiology when ambient levels of pollution measured at distant monitors are used instead of measured exposures

- nonlinear impact models are used, thus making the results unpredictable can cause either an inflation or deflation of estimated effects depending on size of the error
- "regression calibration", i.e. replacing the unmeasured exposures, X by a best predictor
  $\hat{X} = E(x^{unmeasuredvalue} | X^{measuredvalue}, otherdata)$
  is the best general strategy at each time point
- leads to need for space-time models for computing these predictors along with prediction error intervals - hence this course!

Final note: chronic vs acute Morbidity

Acute Effects: Since measurement error effects are moderate, time series methods have made a big industry out of acute health effects analysis by reducing confounder effects.
Chronic effects: Measurement error is large in the study of chronic health effects due to long term exposure.

- Residential history may be incomplete
- Monitoring of environmental hazards quite recent

Hence little done on this topic even though societal costs far greater.

Summary

- Environmental hazards can be hazardous to human health, air pollution being a much studied because its so pervasive
- However, establishing causation runs up against a lot of conceptual and technical issues
- One of these is the need for ways of predicting unmeasured levels and the need for space - time models for doing so

JV Zidek, R White, W Sun, RT Burnett, and ND Le. Imputing unmeasured explanatory variables in environmental epidemiology with application to health impact analysis of air pollution. *Environmental and Ecological Statistics*, 5(2):99–105, 1998.