Roadblocks on the road to causality in environmental epidemiology

Jim Zidek
U BritishColumbia

June 2018

Summary

- Causality: A dilemma for regulators
- Why you should randomize?
- Why you might not want to randomize?
- Why you might change your mind?
- Why you must randomize anyway?
- Now the roadblocks emerge
  - Confounders & the ecologic effect
  - Measurement error
  - Multicolinearity
  - Transfer of causality
- Conclusions

Causality

Hume's 1753 treatise [1] states:

> *"We may define a cause to be an object followed by another, & where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed."*

- **Hume mixes up two definitions of causation** (Stanford Encyclopedia of Philosophy)
- **Second one–counterfactual theory**– developed in succeeding years & supports interpretation of the concentration response function CRF in environmental epidemiology.
- **Elaborate theories of causality now exist** including stochastic versions with applications in health sciences [2]

[1] Hume [2011]

[2] Robins and Greenland [1989], VanderWeele and Robins [2012]

Reproducibility

- Cornerstone of scientific method
- Large number of important scienfific findings using the p–value are nonreproducible [Ioannidis, 2005].
- A p-value ≈ 0.01 does not mean false rejection probability of 1% – its about 11% [3].

[3] [Nuzzo, 2014]

Case study: The regulator's dilemma

JZ on EPA Ozone Air Quality Standards Committee (2005 - 2009)

- Convened to review existing standard (0.08 ppm)
- Process assumes counterfactual causation
- Risk analysis based on:
  - **Concentration response function (CRF):**Observational studies. Regression of morbidity on ambient air pollution levels. Analysis calculates expected decline in morbidity that would be caused by reduction in ozone levels
  - **Exposure response function (ERF):** Randomized exposure chamber studies. Indicates improvement in lung function that would be caused by reduction in ozone from current levels
- **2008 recommendation:** Drop regulatory standard to something in range of $0.060 - -0.070$ ppm. Implemented as 0.070 (ppm) in 2015!

Validity of the counterfactual argument?

COUNTERFACTUAL ARGUMENT: A causes B if B changes when A changes while everything else is held constant.

Validity: How can you hold everything else constant?

- Will reducing ambient ozone levels improve population health by amount predicted by concentration response function CRF?
- Will lung function be improved by amount suggested by the exposure response function ERF?
- What about multi–collinearity?

Statistical road to causality – randomized designs

Assign experimental treatments at random to experimental units!

- **Manufacturer's dilemma!**
- **Fisher goes to Rothamsted (1990):** Argued for randomized design based on spatial fertility gradients in the soil.

- **Salk polio vaccine trial:** Randomized case–control clinical trials in the 1950s–famous experiment that enshrined randomized trials in medical research
- **US lakes survey (2009):** Stratified survey sample – water samples kriged over land!

Variable regional randomization protocols?

SIMPSON'S PARADOX: Fictitious example. Involves data from Regions Q & R aggregated as one P.

	Treatment T or Control $\bar{T}$	
Result	T	$\bar{T}$
Died	273 (78%)	289 (83%)
Survived	77	61

CONCLUSION: **Control $\bar{T}$ is better!**

Disaggregated data for Regions Q (Top) & R (Bottom).
Treatment wins in both regions!

	Treatment T vs Control $\bar{T}$	
Result (1000s)	T	$\bar{T}$
Died	81 (93%)	234 (87%)
Survived	6	36

	Treatment T or Control $\bar{T}$	
Result (1000s)	T	$\bar{T}$
Died	192 (73%)	55 (69%)
Survived	71	25

Defining a Simpson–disaggregation

Given two regions Q, R and

Treatment vs Control: T & $\bar{T}$

Success vs Failure: S & $\bar{S}$

Confounder at two levels: C & $\bar{C}$.

Let

$$\begin{aligned} P_1 &= P(S \mid T)^{high\ prob} \\ P_2 &= P(S \mid \bar{T})^{low\ prob} \\ \alpha_1 &= P(\bar{C} \mid T) \\ \alpha_2 &= P(\bar{C} \mid \bar{T}) \end{aligned}$$

A Simpson–disaggregation for $0 \leq P_2 < P_1 \leq 1$ defined as:

$$P_i = \alpha_i Q_i + (1 - \alpha_i) R_i,\ i = 1, 2$$

with $0 \leq Q_1 \leq Q_2 \leq$ & $0 \leq R_1 \leq R_2 \leq 1$,

A Simpson–disaggregation can exist if & only if [4]

$$\Gamma < \frac{P_2/(1-P_2)}{P_1/(1-P_1)}$$

where Γ involves the critical ratio $\theta = P(C)/[P(\bar{C})]$.

Research questions:

- How can this result be extended to a multiplicity of regions?
- Can the result be reformulated in the context of causal statistical inference?
- The ecological effect?

[4] Zidek [1984]

Some people don't like randomization!

- **Designs for monitoring hazardous environmental fields:** Often model based. E.G. entropy approach yields maximally "informative" designs. Based on an information criterion for stochastic models of the random field, e.g ozone.
- **Bayesians:** Use nonrandomized designs due to likelihood principle

Case study: The ECMO trials

Ware [1989]: Describes **ECMO randomized clinical trials**. ECMO oxygenates blood of infants using cardiopulmonary bypass technology. Ware et al conducted randomized trials at Harvard for infants with severe hypertension (low blood flow). But previous data were available. A play–the–winner strategy was used for treatment allocationl.

Don Berry (Bayesian discussant): **criticizes trial.** Calls it unethical. Shows how Bayesian methods could have been used for a model-based analysis. Uses historical data. **Concludes:**Randomized trials not perfect inferential tools & “ethical issues dictate their demise”

But what if many Bayesians involved?

Group Bayesian i.e. multi-agent decision analysis complex. Involves competing utility functions & prior knowledge bases [5]

Example: Two Bayesians B_1 & B_2: Must produce common estimator of multivariate normal mean θ.
B_i**'s posterior:**

$$\pi_i(\theta) \propto \exp\{-\frac{1}{2}(\theta - \theta_i)^T \Sigma_i (\theta - \theta_i)\}.$$

B_i**'s conjugate utility:**

$$u_i(\hat{\theta} \mid \theta) \propto \exp\{-\frac{1}{2}(\hat{\theta} - \theta)^T \Omega_i^{-1} (\hat{\theta} - \theta)\}$$

[5] Weerahandi and Zidek [1983]

Expected utility for nonrandomized estimation rules [6]:

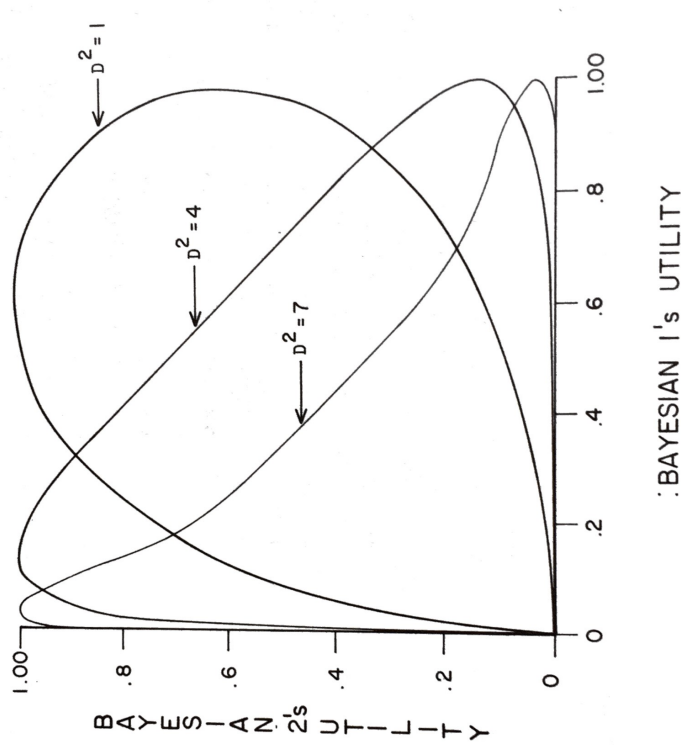
$$B_i(\hat{\theta}) \propto \exp\{-\frac{1}{2}(\hat{\theta}-\theta_i)^T(\Omega_i+\Sigma_i)^{-1}(\hat{\theta}-\theta_i)\}$$

Definition: distance D between B_1 & B_2:

$$D^2 = (\theta_1-\theta_2)^T(\Omega_i+\Sigma_i)^{-1}(\theta_1-\theta_2)$$

[6] Weerahandi and Zidek [1983]

Plot of $(B_1(\hat{\theta}), B_2(\hat{\theta}))$ for varying $\hat{\theta}$ & D:



Note: Coin toss optimal when $D^2 < 4$

Research question:

- Design a multi–agent clinical trial.

Randomized double blind trials?

"SELECTIVE TRIALS: A PRINCIPAL-AGENT APPROACH TO RANDOMIZED CONTROLLED EXPERIMENTS" [7]

Paper assumes both Bayesian Experimenter & Bayesian Subjects. Result: randomized trial. By CERC Economics faculty member at UBC & co–authors.

[7] Chassang et al. [2012]

But sometimes randomization just not on!

- **Unethical.** Can't force grad students into exposure chambers to estimate ozone dose response models
- **Can yield dissimilar samples with moderate sample sizes:** Can align with confounders by chance.
- **Often practical only with small samples to ensure controlled conditions:** May be impossible in social science e.g. economics.
  - Strawberries grow well in ozone infused tents on campus–but what about in vast fields in all weather conditions?

Observational studies may be inevitable**********

But good design principles still apply.

- **Experimental units from high & low risk spatial domains"** Ensures contrasts & statistical power
- **Case-crossover designs:**Subjects serve as own controls (compare affected group on high ozone days with that same group on low ozone days
- **Longitudinal data analysis/designs** [8]**:** Spatial confounders eliminated while providing strong contrasts.

[8] Liang and Zeger [1986]

Find “smoking gun” observational studies**********

Example – MESA study: Designed to test hypothesis:

> *Inhaled PM (airborne particulates) initiate production by lung cells of inflammatory mediators & oxidants that become blood borne & target the cardiovascular system.*

Most heavily funded study (MESA) in history on health effects of air pollution

- US EPA funded ten-year study – began in 2000
- Involved 1000’s of participants
- Multi–ethnic study of Atherosclerosis (MESA) Air Pollution
- Represented diverse geographical areas
- Centered at the University of Washington but subjects from New York, Maryland, North Carolina, Minnesota, Illinois, & California
- Examined air pollution exposures – progression of cardiovascular disease over time relationship.

Conclusions??

Alternative: Bradford–Hill route.

Sir Austin Bradford Hill. English epidemiologist:

Strength of association: The stronger the association between a risk factor & outcome, the more likely the relationship is to be causal.

Specificity: There must be a one–to–one relationship between cause & outcome.

Biological plausibility: There should a plausible biological mechanism which explains why exposures might cause adverse health effects.

Biological gradient: Change in disease rates should follow from corresponding changes in exposure

Temporal gradient: Exposure must precede outcome.

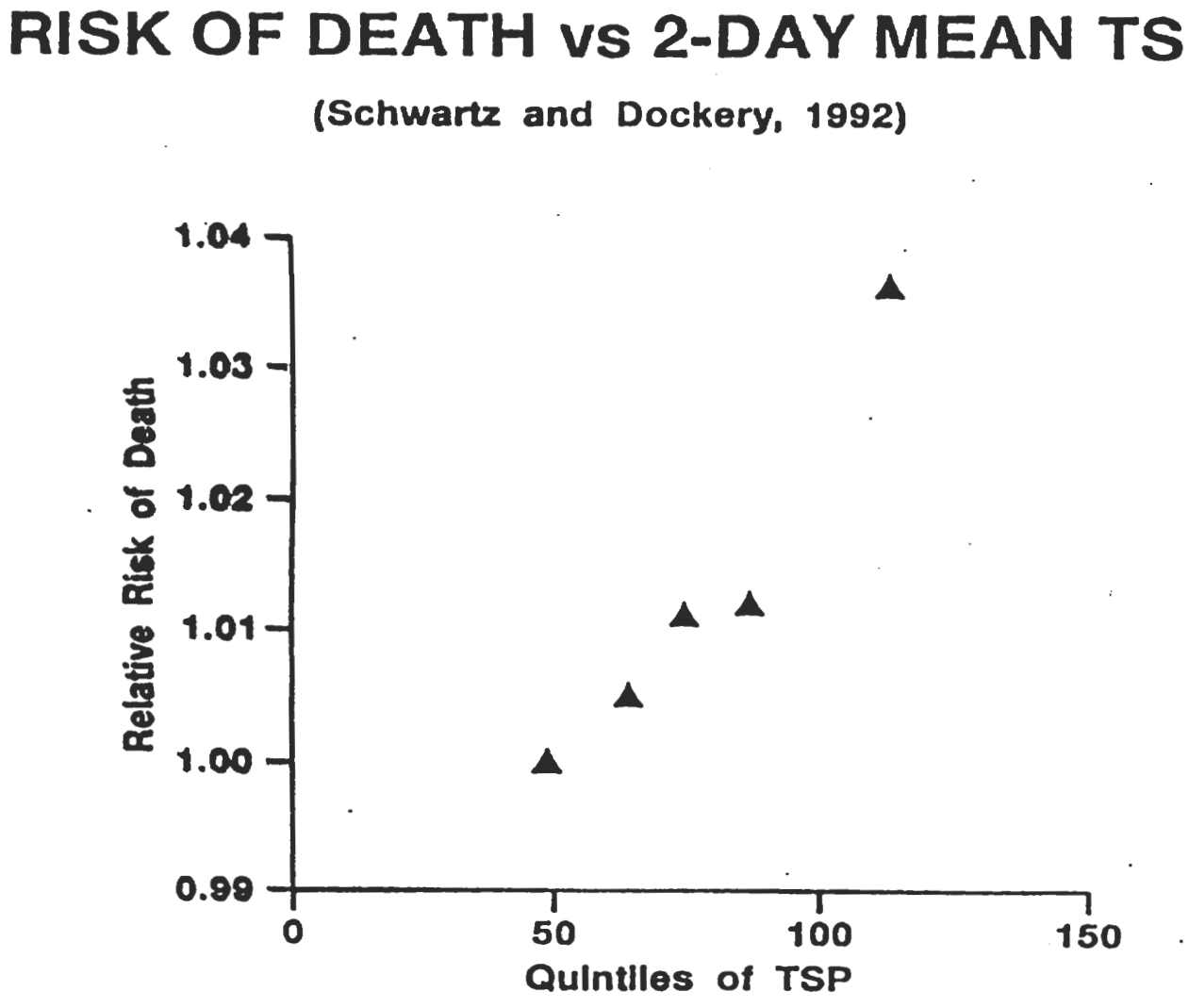
Consistency of results: The same effect should be seen among different populations, when using different study designs & at different times.

Coherence: Does the relationship agree with the current knowledge of the natural history/biology of the disease?

Experimental evidence: Is there evidence that removal of the exposure alters the frequency of the outcome?

Analogy: Similar findings in another context may support the claim of causality in this setting.

Biological gradient: Total suspended particulates*****



Temporal gradient: The famous 1952 London fog******

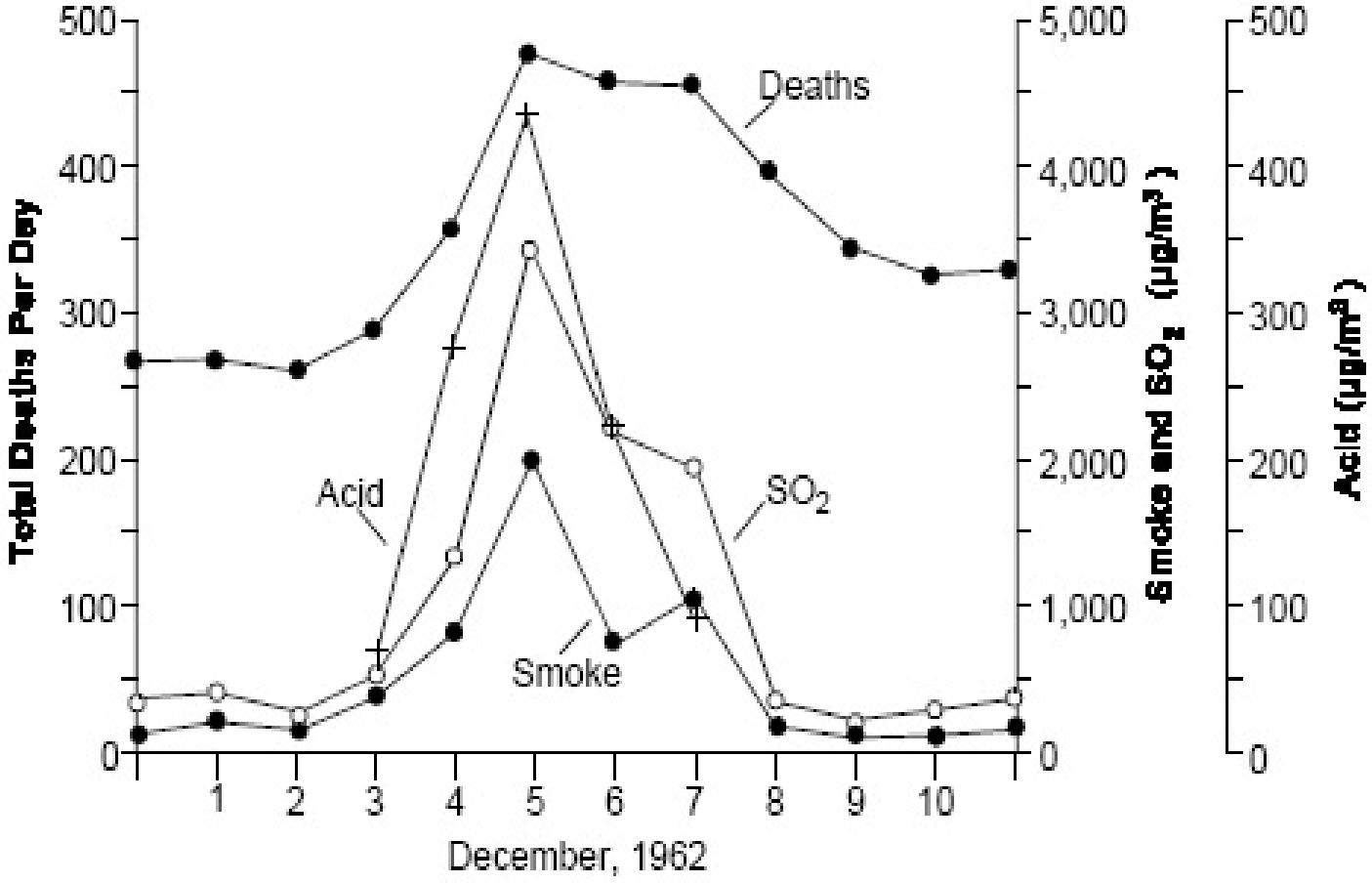


Figure 12-10. December 1962, London pollution episode.

Consistency of results: Across various studies

Relative risk of respiratory health outcome vs particulate increase [9]

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%

[9] GOLDBERG [1996]

Challenging conditions on the road

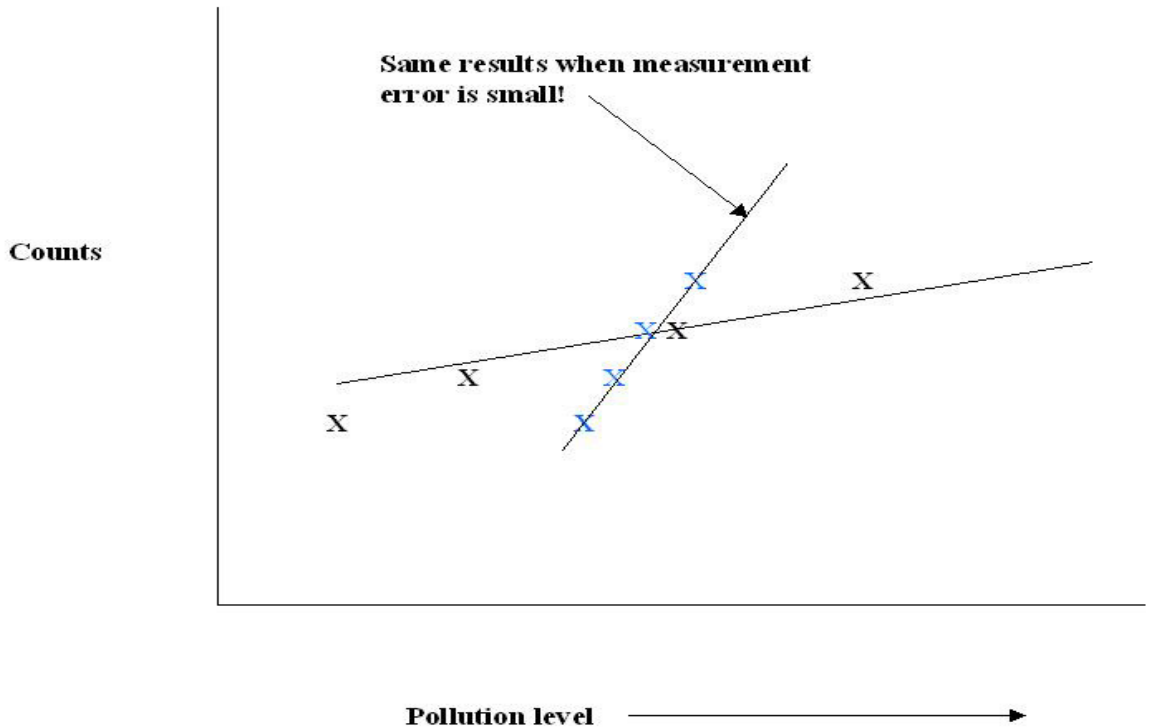
- **Effects (relative risks) may be small e.g 2%:** Need sophisticated spatio–temporal modeling. Causation signals difficult to detect in the noise.
- **Measurement errors can be large & of two types:** Classical & Berkson
- **Extreme value effects difficult to detect:** No implementable theory for large spatial domains.

Specific challenges #1: Measurement error

Measurement error:

- *Can have serious & unpredictable effects:* Generally thought to shrink the relative risk toward null hypothesis of no association.
- *Nonlinear models can go either way:* shrink or expand RRs!

*****


Same results when measurement error is small!
Counts
Pollution level

Specific challenges #2: Collinearity

A major problem in environmental epidemiology– discussed later.

- Possible that everyone of a suite of pollutants can be found to be individually significant, yet none significant when all fitted together.
- difficult to unravel say temperature and ozone that are highly correlated & both a risk factor for mortality.
- **What is the true cause of death?**

Example of collinearity*****

Daily **hospital admissions** for **respiratory morbidity** in census subdivisions for southern Ontario vs O_3 & SO_4. They competed as predictors depending on lags. Next table reveals the competition in stepwise fitting.

SO_4 **= sulfate.** $Temp$ **= temperature.“L” = 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	$O_3\ L0$					

Challenges combined: Measurement error & collinearity

Transfer of Causality:
$[Y \mid x] = \text{Poisson}(\exp \alpha_0 + \alpha_1 x)$, $x = \text{hazard}$, $w = \text{covariate}$, measured as X, W with correlation ρ & $\sigma_X^2 >> \sigma_W^2$.

- **Fitted model:** $a_0 + a_1 X + a_2 W$.
- **Result:** a_1 *nonsignificant* but a_2 *significant* when $\rho = 0.9$, $\sigma_X^2 > 0.5$ & $\sigma_W^2 \approx 0$, for example.

Conclusion: Although x causes Y, W is identified instead. Causality transferred! [10]

[10] Zidek et al. [1996]

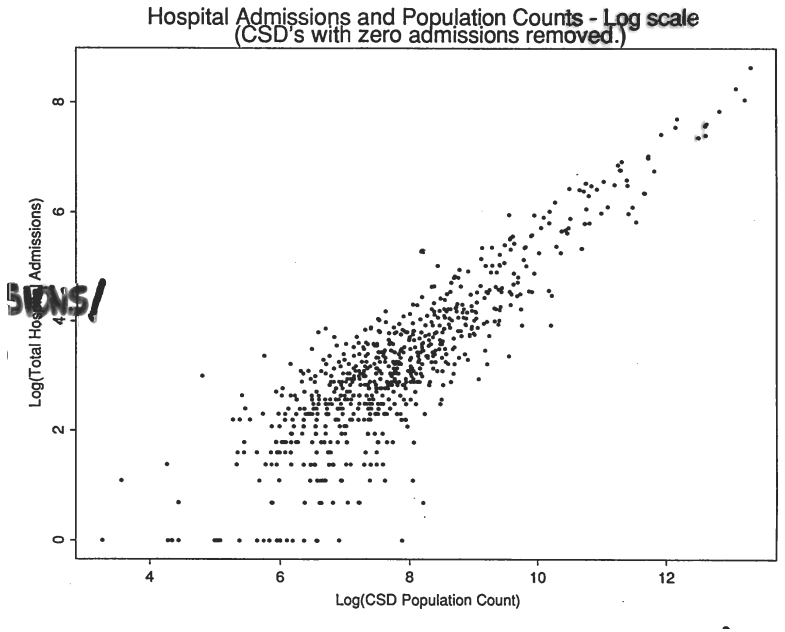
Specific challenges #3:Confounders

Given: Environmental hazard: Does it cause morbidity or mortality?

Confounders (observed & unobserved) may spuriously point to causation. E.g. I**hi ozone concentrations** & **mortality** association could be due to **high temperatures**.

Example*******

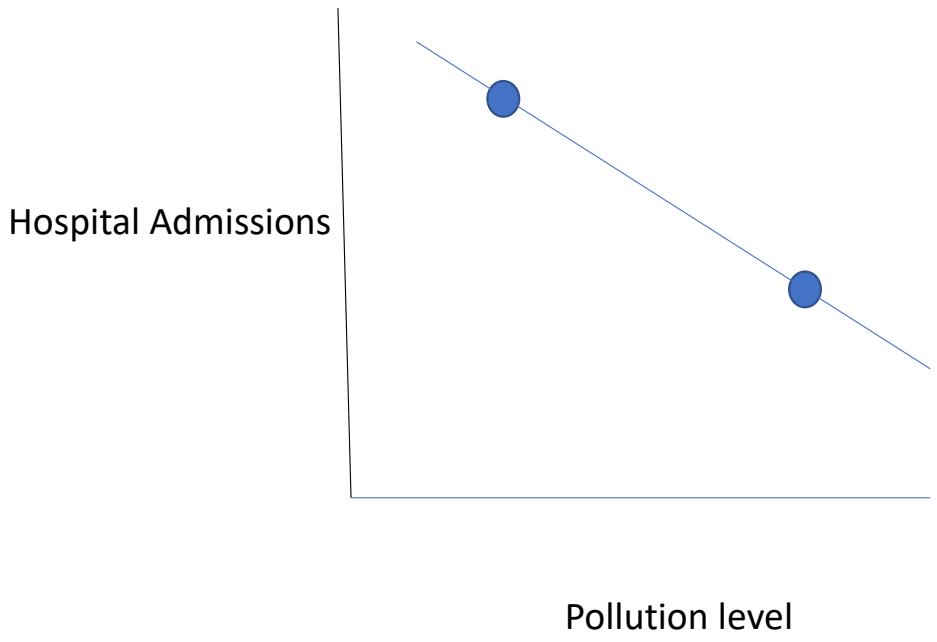
Log(Hospital – admissions counts) vs Log (ozone concentrations) for census subdivisions in Ontario. Inference of cause seems valid. But likely due to **confounder** populations size.



Specific issue #4: Ecological measurement error bias

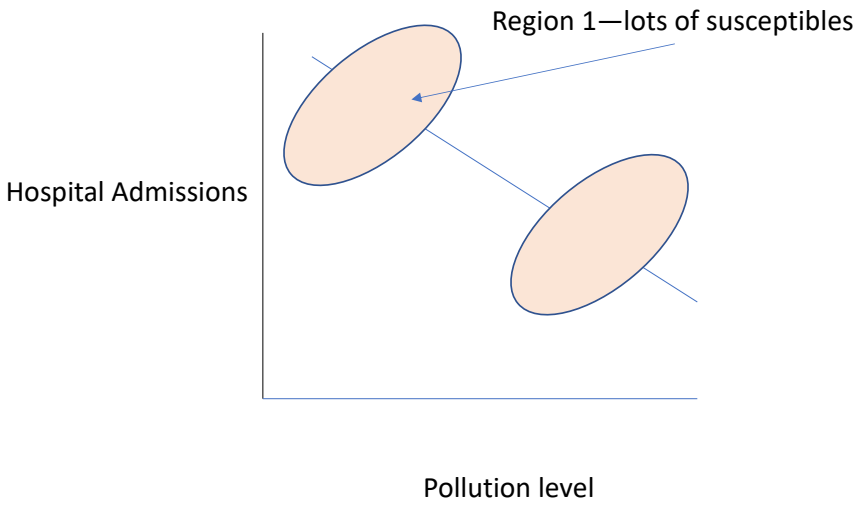
Aggregate data: Can give misleading picture of individual effects. **Ecological fallacy:**Extension of *Simpson's paradox*.

Hypothetical example. Figure suggests pollution is good thing.



The real story!

Region 2's population is just less susceptible than Region 1's



Research question:

- Can concept of **Simpson disaggregation** be extended to embrace the **Ecological fallacy**

Dealing with #1 Measurement error*****.

Arises in environmental epidemiology when ambient pollution concentrations used as surrogate for measured exposures

- The **effect is unpredictable for nonlinear health impact models**– can either inflate or deflate estimated effect depending on size of the error
- **Regression calibration**, i.e. replacing the unmeasured exposures, X by a best predictor $\hat{X} = E(x^{unmeasuredvalue}|X^{measuredvalue}, otherdata)$ at each time point best general strategy.
- **Result:** need for spatio–time models for computing the predictors & their prediction error intervals

Spatial prediction****

A general approach:

- **Basic building blocks:** Uncorrelated clusters i
- **Health outcomes (eg deaths) in each cluster:** $\{Y_{it}\}$, for time point t (e.g. day), & cluster i
- **Pollution concentration (& covariate) vectors:** $\{X_{it} = (X_{it1}, \ldots, X_{itk})\}$–may be hi–pass filtered to unmask blip effects
- **A health effects model:**

$$E[Y_{it} \mid X_{it}, \mathbf{a_i}] = m_{it} \exp\left(\mathbf{a}_i^T X_{it}\right)$$

  - with m_{it} a fixed factor accounting for population size, day of week & low frequency seasonal components

Dealing with # 3 Confounders*****

- **Cross sectional studies:** Look at different samples of primary units, eg census subdivisions at successive times.
- **Longitudinal studies:** Will keep the same sample & make repeated measures.
  - Measure long term trends better.
  - Allow unit changes in health outcomes to be compared directly against pollution changes – reduces possible intervention by confounders!
- **Modelling the effects:**
  - *Confounders* can be included in the response regression model.
  - Can be used to re–weight data using inverse of propensity scores
  - or both!

Dealing with # 3 Confounders

Propensity scores: Can help contend with bias in the quest for "cause". Defined by:

$$p_h = P(Z_h^{membership\ in\ exposure\ group} = 1 \mid X_h^{observed\ confounder})$$

Vast literature. See for example Stephens et al (2016) [11] Used in four different ways:

1. matching;
2. stratifying;
3. adjusting;
4. inverse probability weighted.

[11] Graham et al. [2016]

Inverse probability re–weighting

Let:

$$w_h = Z_h p_h^{-1} + (1 - Z_h)(1 - p_h)^{-1})$$

Replace Y_h by $w_h Y_h$. boost the Ys that were under sampled due to bias e.g. in

$$\sum_{h=1}^{N} Y_h$$

Sometimes called the **Horwitz–Thompson estimator**.

Modelling ecological measurement bias

Recent approach[12]**:** Builds on Wakefield & Shaddick (2005) [13].

For Subject $i \in A_k^{\text{Region}}$, assume $Y_{ki}^{\text{Response}} \sim Bernoulli(p_{ki})$,

$$log(p_{ki}) = \beta_0 + \beta_1 X_{ki}^{Exposure}.$$

Then for A_k

$$Y_k^{Region\ total} = \sum_{i=1}^{N_k} Y_{ki}$$

with

$$E(Y_k) = N_k q_k,\ q_k = N_k^{-1} \sum_{i=1}^{N_k} \exp\left(\beta_0 + \beta_1) X_{ki}\right)$$

[12] Wang et al. [2017]

[13] Wakefield and Shaddick [2006]

But individual responses in A_k unknown! So average number of responses is stochastic

$$\int_{s \in A_k} i(s) f_k(s) ds.$$

Here $f_k = \text{population density}$ and
$i(s) = \text{conceptual incidence surface}$. Let $f_k \equiv 1$ for simplicity and

$$logit(p(s) = \beta_0 + \beta_1 X_k(s)^{GP} + \eta_k^{CAR\ process}$$

Reflects a general approach: Use latent processes within areas to reflect unobserved individual responses to reflect variability/bias not captured by standard ecological model.

Wrap up: So what do we say to the regulators?

- How should they determine regulatory standards e.g 0.060ppm?
- How would they know if estimated risk reductions are attained?
- Should randomization be abandoned?
- What do we now know about measurement error? Confounders? Ecological measurement bias?
- In general is the theory for causal inference converging?
- What about a Conference/workshop dedicated to causal inference for regulatory policy?

Title: What Role Should Statistics & Statisticians Play in Environmental Policy & Regulation?

Abstract: Environmental statistics is essential to protecting human health & the natural environment. Statistical verbiage & references to methodology frequently appear in policy & guidance documents, as well as rules & regulations, suggesting or even mandating specific, & sometimes inappropriate, statistical methods. Method specificity is motivated & constrained by the requirement a non-statistician have access to simple software packages that produce results automatically. This conflicts with the fundamental paradigm of statistical science that rules out such uncritical, mechanical use of methods, often driven by flow charts. Instead, defensible conclusions often require context-specific analyses to support decision-making based on a wide variety of data & designs for data collection. **The field of statistics continues to evolve, & discourse on the interface of policy, regulation, & statistics is key to addressing the complex environmental issues we face.** This panel brings together individuals with different experiences & perspectives to introduce fundamental issues, generate discussion, & motivate future participation.

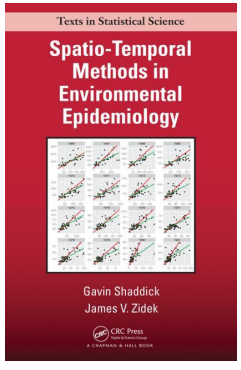
Concluding slide

Contact Information:

James V Zidek
email: jimstat.ubc.ca
home page: http://www.stat.ubc.ca/ jim/

Reference:

Spatio-Temporal Methods in Environmental Epidemiology.
Gavin Shaddick & James V. Zidek
Chapman & Hall/CRC



Sylvain Chassang, Padró I Miquel, Erik Snowberg, et al. Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102(4): 1279–1309, 2012.

MARK S GOLDBERG. Particulate air pollution and daily mortality: who is at risk? *Journal of aerosol medicine*, 9(1): 43–53, 1996.

Daniel J Graham, Emma J McCoy, David A Stephens, et al. Approximate bayesian inference for doubly robust estimation. *Bayesian Analysis*, 11(1):47–69, 2016.

David Hume. *An enquiry concerning human understanding*. Broadview Press, 2011.

John PA Ioannidis. Contradicted and initially stronger effects in highly cited clinical research. *Jama*, 294(2):218–228, 2005.

Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Regina Nuzzo. Scientific method: statistical errors. *Nature News*, 506(7487):150, 2014.

James Robins and Sander Greenland. The probability of causation under a stochastic model for individual risk. *Biometrics*, pages 1125–1138, 1989.

Tyler J VanderWeele and James M Robins. Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica*, 22(1):379, 2012.

Jon Wakefield and Gavin Shaddick. Health-exposure modeling and the ecological fallacy. *Biostatistics*, 7(3):438–455, 2006. ISSN 1465-4644. doi: 10.1093/biostatistics/kxj017.

Feifei Wang, Jian Wang, Alan Gelfand, and Fan Li. Accommodating the ecological fallacy in disease mapping in the absence of individual exposures. *Statistics in medicine*, 36(30):4930–4942, 2017.

James H Ware. Investigating therapies of potentially great benefit: Ecmo. *Statistical science*, pages 298–306, 1989.

S Weerahandi and JV Zidek. Elements of multi-bayesian decision theory. *The Annals of Statistics*, pages 1032–1046, 1983.

JAMES V Zidek, HUBERT Wong, ND Le, and RICK Burnett. Causality, measurement error and multicollinearity in epidemiology. *Environmetrics*, 7(4):441–451, 1996.

Jim Zidek. Maximal simpson-disaggregations of 2$\times$ 2 tables. *Biometrika*, 71(1):187–190, 1984.