Introduction to Causal Inference

Lan Liu

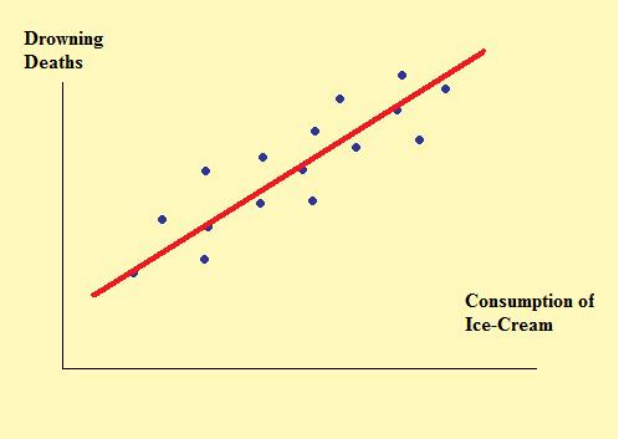
University of Minnesota at Twin Cities

liux3771@umn.edu

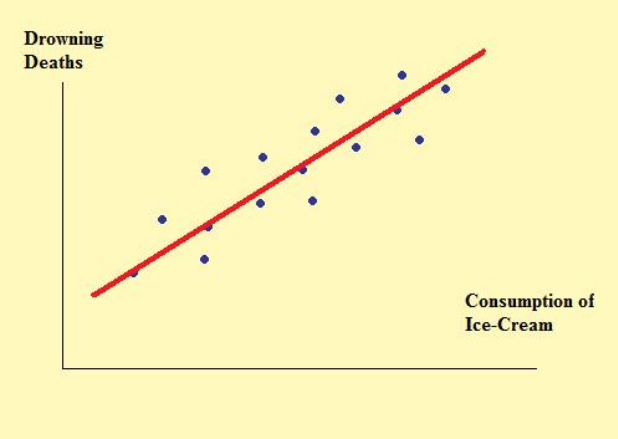
Table of contents

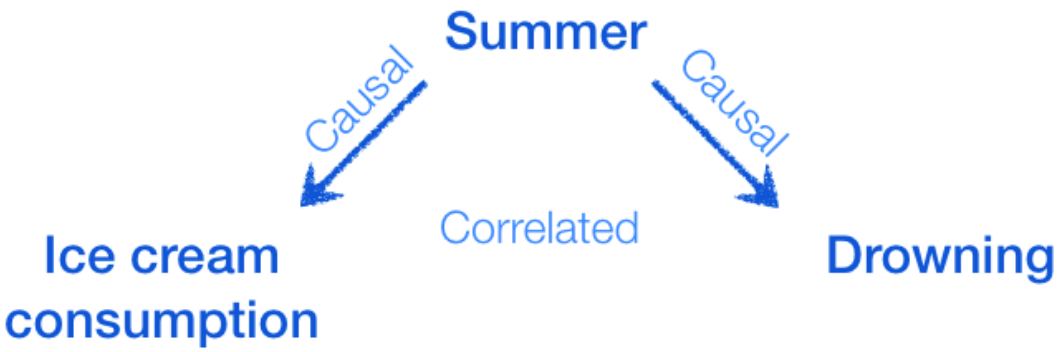
The Danger of Ice Cream

The Danger of Ice Cream

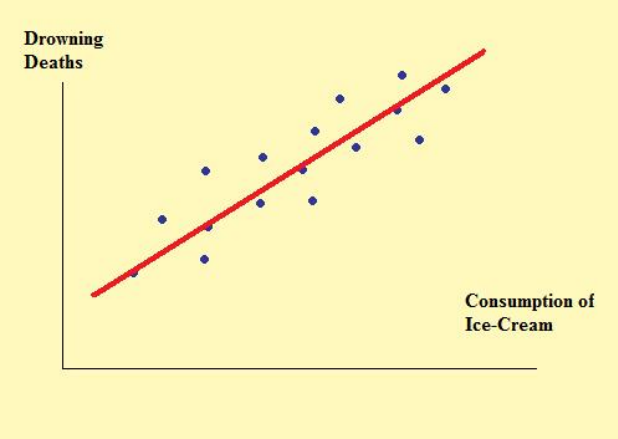


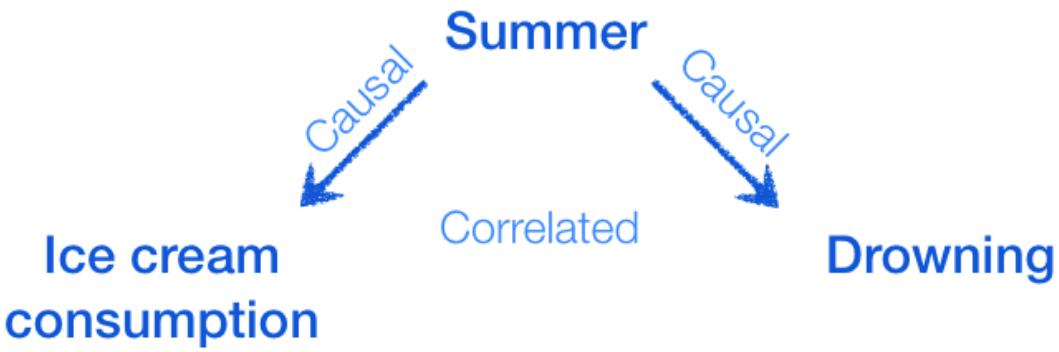
The Danger of Ice Cream





The Danger of Ice Cream





- "Confounding Bias"

Marriage vs Longivity

Science points to a very easy way to be happier, have less stress, reduce your risk of dying from cancer and heart disease, and potentially live longer:

Marriage vs Longivity

Science points to a very easy way to be happier, have less stress, reduce your risk of dying from cancer and heart disease, and potentially live longer: Simply get married!!

Marriage vs Longivity

Science points to a very easy way to be happier, have less stress, reduce your risk of dying from cancer and heart disease, and potentially live longer: Simply get married!!

- "Reverse Causality"

World War II

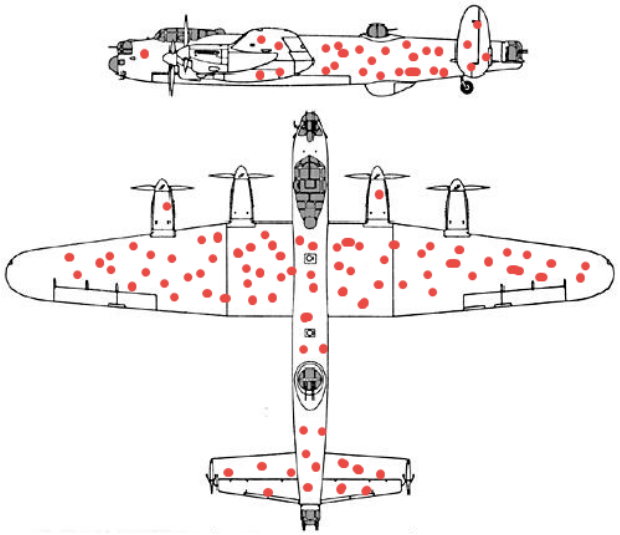
Abraham Wald (THE Wald as in Wald test)

- Britian vs Germany
- Bomber: cumbersome, easily hit by fighters
- Install armour: heavy
- Look at aircraft that had returned from missions
  - add to the most hitted areas

World War II

Abraham Wald (THE Wald as in Wald test)

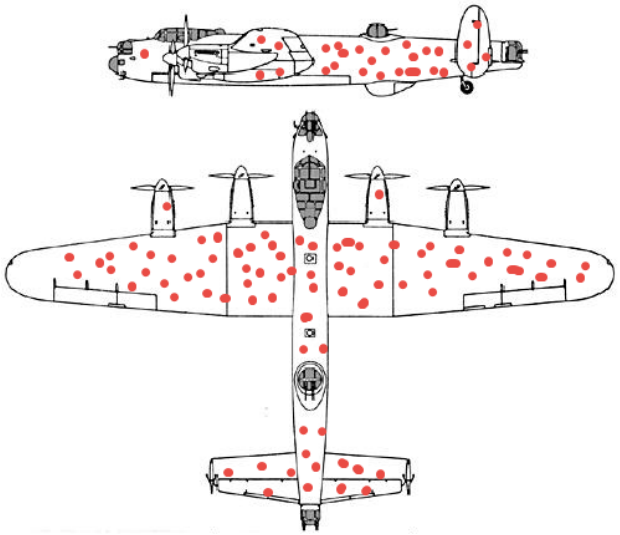
- Britian vs Germany
- Bomber: cumbersome, easily hit by fighters
- Install armour: heavy
- Look at aircraft that had returned from missions
  - add to the most hitted areas



World War II

Abraham Wald (THE Wald as in Wald test)

- Britian vs Germany
- Bomber: cumbersome, easily hit by fighters
- Install armour: heavy
- Look at aircraft that had returned from missions
  - add to the most hitted areas



- "Selection Bias"

How to Make Causal Inference

How to Make Causal Inference

- Time machine ...

How to Make Causal Inference

- Time machine ...
- Parallel universe
  - Potential outcomes: Y_0, Y_1.
  - Individual causal effect $Y_1 - Y_0$

Table 2.1

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Cyclope	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

  - Movies: Sliding Door, Mr. Nobody

How to Make Causal Inference

Key: have control over intervention

Golden rule: randomization

Not so easy to randomize ...

- Randomization may be costly!
  - E.g., google search story, try search: BMW, sun country, iphone

Not so easy to randomize ...

- Randomization may be costly!
  - E.g., google search story, try search: BMW, sun country, iphone

Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment

Thomas Blake, Chris Nosko, Steven Tadelis

First published: January 2015 Full publication history

DOI: 10.3982/ECTA12423 View/save citation

Cited by: 15 articles Citation tools

Am score 2

View issue TOC
Volume 83, Issue 1
January 2015
Pages 155–174

Abstract

Internet advertising has been the fastest growing advertising channel in recent years, with paid search ads comprising the bulk of this revenue. We present results from a series of large-scale field experiments done at eBay that were designed to measure the causal effectiveness of paid search ads. Because search clicks and purchase intent are correlated, we show that returns from paid search are a fraction of non-experimental estimates. As an extreme case, we show that brand keyword ads have no measurable short-term benefits. For non-brand keywords, we find that new and infrequent users are positively influenced by ads but that more frequent users whose purchasing behavior is not influenced by ads account for most of the advertising expenses, resulting in average returns that are negative.

Not so easy to randomize....

- People don't listen....
  - E.g., non-compliance $\rightarrow$ smaller treatment effect
  - Confounding
- Ethical reasons: e.g., smoking vs lung cancer

Measured confounding

- E.g., Study: working out vs body fat
  - Subject matter knowledge: women differ from men!

Measured confounding

- E.g., Study: working out vs body fat
  - Subject matter knowledge: women differ from men!
    - woman: gym goers vs non goers
    - man: gym goers vs non goers
    - Stratify on gender

Measured confounding

- E.g., Study: working out vs body fat
  - Subject matter knowledge: women differ from men!
    - woman: gym goers vs non goers
    - man: gym goers vs non goers
    - Stratify on gender
  - Better knowledge: not only gender, but also age, race, eating habits matter!

Measured confounding

- E.g., Study: working out vs body fat
  - Subject matter knowledge: women differ from men!
    - woman: gym goers vs non goers
    - man: gym goers vs non goers
    - Stratify on gender
  - Better knowledge: not only gender, but also age, race, eating habits matter!
  - Even better knowledge: what if genes also matter?!

Measured confounding

- E.g., Study: working out vs body fat
  - Subject matter knowledge: women differ from men!
    - woman: gym goers vs non goers
    - man: gym goers vs non goers
    - Stratify on gender
  - Better knowledge: not only gender, but also age, race, eating habits matter!
  - Even better knowledge: what if genes also matter?!
- Only need to stratify on the value of propensity score, i.e., $\Pr(\text{go to gym}|X) \rightarrow$ propensity score matching

Unmeasured confounding

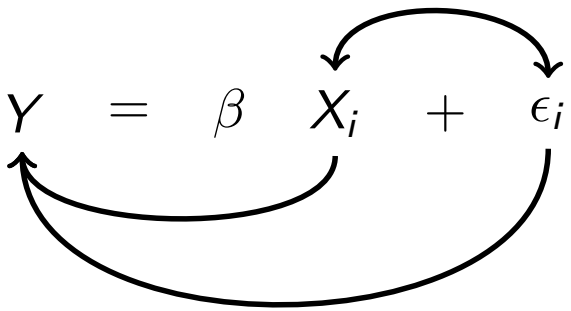


Figure: Causal diagram of the confounding bias

- $$\hat{\beta}^{LS} = \frac{\Delta y}{\Delta x} = \frac{\Delta y_x + \Delta y_\epsilon}{\Delta x} = \beta + \frac{\Delta y_\epsilon}{\Delta x}$$

- Biased!

Unmeasured confounding

$$\begin{tikzcd} Y \arrow[rr, bend left] & = & \beta & X_i \arrow[l, bend left] & + & \epsilon_i \\ & & & Z_i \arrow[u] & & \end{tikzcd}$$

Figure: Causal diagram of the confounding bias

- One solution: Instrumental variable $\rightarrow$ Unbiased!

$$\hat{\beta}^{IV} = \frac{\Delta y}{\Delta x} = \frac{\Delta y_x}{\Delta x} = \beta$$

Mediation

- Mediation: causal pathway, underlying mechanism
  - E.g.,

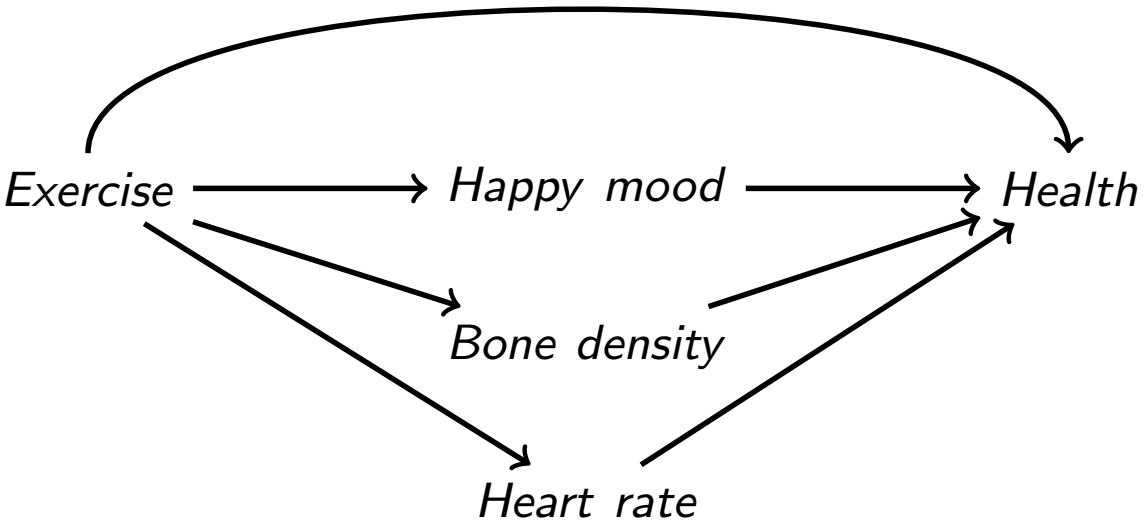
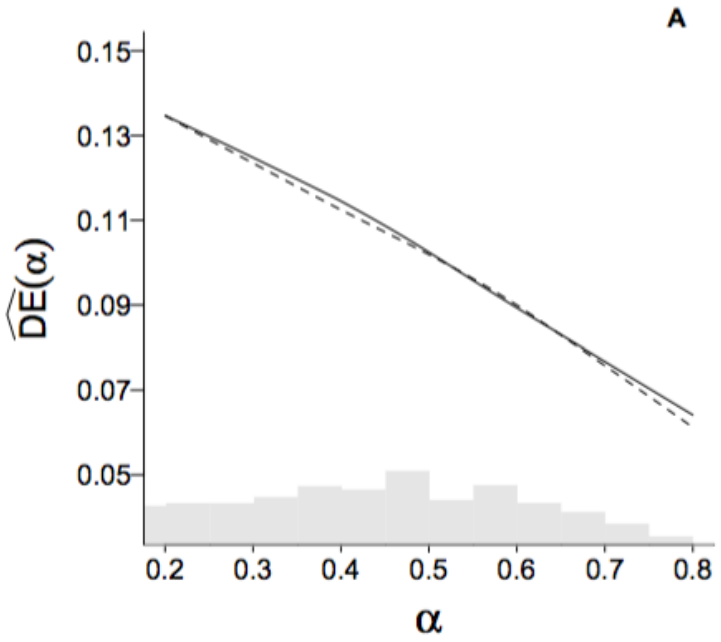


Figure: Causal diagram of the causal pathways from exercise to health

Interference

- Interference: your outcome also depends on other people's treatment
  - E.g., flu vaccine study → herd immunity



Topics in Causal Inference

Other topics includes:

- measurement error (surrogate)
- heterogeneity treatment effect
- graphical models
- ...

Tools we use...

Tools we use...

Almost everything in statistics ...

- Multiple comparison
- Hypothesis testing
- Parametric modeling
- Semiparametric efficiency
- Nonparametric smoothing
- Structural modeling
- ...

Causal inference is a special type of statistics, where we care only certain type of association, which is due to causation ...

Do Industry ppl care?

Do Industry ppl care?

Of course!

- Tech companies: e.g., facebook (interference), amazon, bing (causal effect of advertisement)...
- Insurance companies: effect of training program for sales persons
- Finance: policy (e.g., increase interest rate) consequence
- Pharmaceutical companies: curing ppl, who are we curing ...
- Sports: effect of certain play strategy
- ...

My recent research

Optimal Criteria to Exclude the Surrogate Paradox

Introduction

- What is surrogate?

Introduction

- What is surrogate?

Scapegoat

Introduction

- In biomedical and econometric studies, the measurement of the primary endpoint may be
  - expensive
  - inconvenient
  - infeasible to collect in a practical length of time.
- Surrogate variables/ biomarkers are usually used as substitutes for the primary outcomes.
  - In cancer studies, the primary outcome is death;
  - Surrogate: tumor shrinkage/ other laboratory measure $\rightarrow$ reduce the cost or the duration of the clinical trials

Horrible Consequences

- Eg 1., Lipid levels (especially total cholesterol levels) $\rightarrow$ predictors of cardiovascular-related mortality.
  - However, the use of cholesterol-lowering agents $\rightarrow$ increase in overall mortality (Gordon, 1995).
- Eg 2., Anti-arrythmia drug Tamnbocor $\rightarrow$ suppresses arrythmia $\rightarrow$ death of over 50,000 people!!

Surrogate paradox

- The surrogate paradox: $+$ treatment effect on the surrogate, $+$ surrogate effect on the true endpoint $\Rightarrow -$ treatment effect on the true endpoint.
- Even the sign of the treatment effect is hard to predict, not to say magnitude!!!
- Happen even in randomize studies

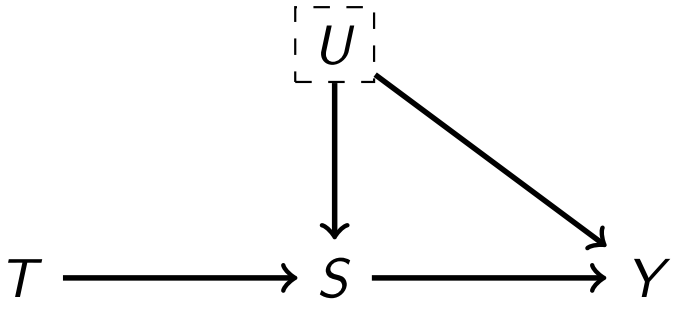


Figure: Causal diagram of the strong surrogate S for the effect of the treatment T on outcome Y.

Methods

- Long story short: old methods all assume unverifiable assumptions, thus may not be practical to use
- We developed bounds for the treatment effect with surrogate without any unverifiable assumptions
  - We used linear programming to solve this
  - We show that it is not enough to avoid the surrogate paradox merely with the ACE of surrogate on outcome being positive, instead, we require its magnitude to pass certain positive threshold.
  - Transportability; testability; optimality.

Excluding the Paradox

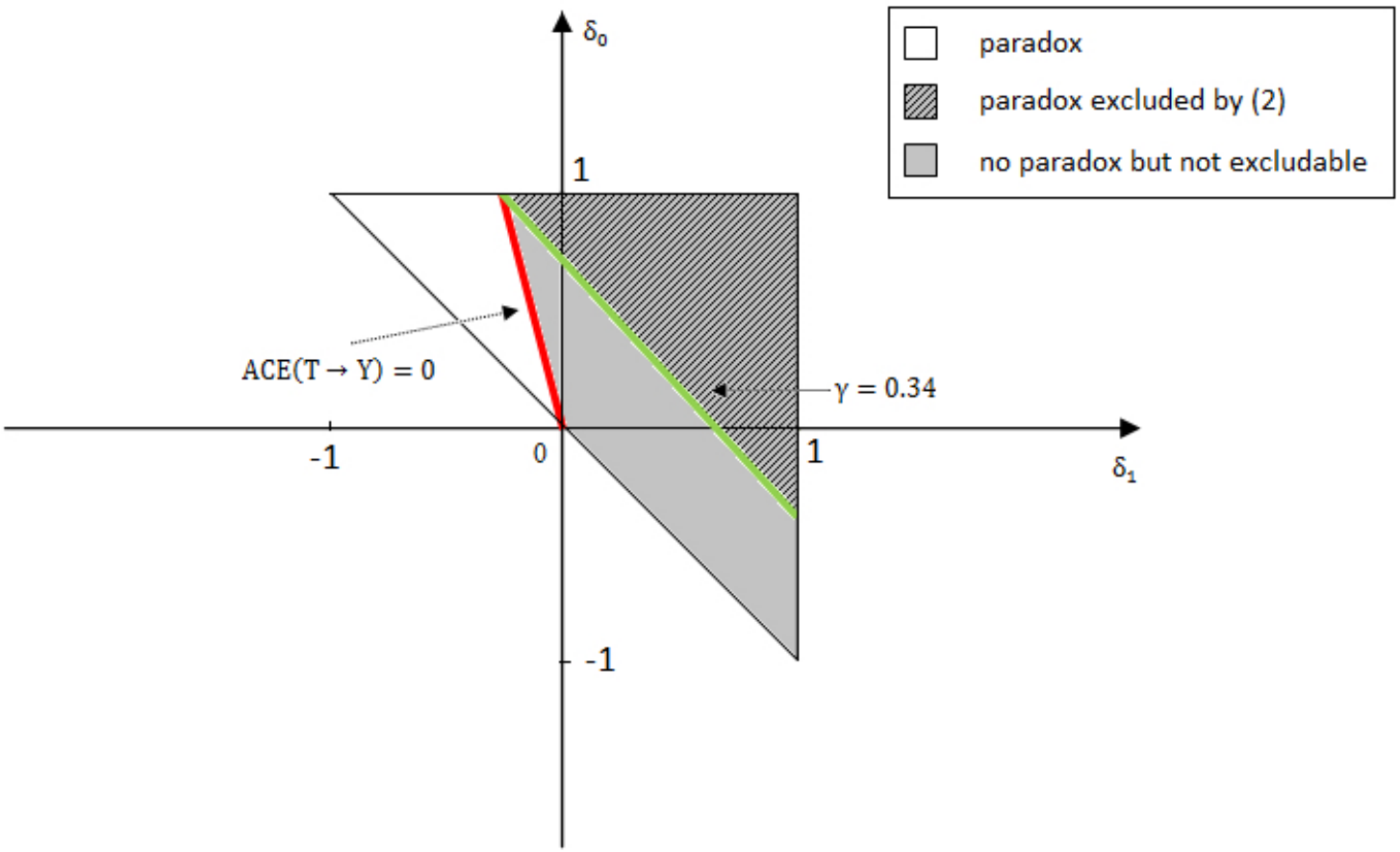


Figure: Partition of the parameter space of (δ_0, δ_1)

Statistical Analysis

Anti-hypertension Drugs

- Thus, we conclude that for evaluating the effect of anti-hypertension drug on the long-term death, using high blood pressure as a surrogate cannot guarantee the bounds to exclude null.
- That is, if the unmeasured confounders have certain value, it is possible that the treatment has a possible effect in reducing the high blood pressure and lowering the high blood pressure could reduce the death rate, but the treatment could increase the death rate.
- Thus, for the development of such anti-hypertension drug, it is recommended to also collect the information on the long-term death rate.

Entry Level Causal References

- Book by Hernan and Robins https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- Imbens, Guido W., and Donald B. Rubin. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, 2015.

Edge cutting References

H. Chen, Z. Geng, and J. Jia.
Criteria for surrogate end points.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69:919–932, 2007.

M.G. Hudgens and M.E. Halloran.
Causal vaccine effects on binary postinfection outcomes.
Journal of the American Statistical Association, 101:51–64, 2006.

C. Ju and Z. Geng.
Criteria for surrogate end points based on causal distributions.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72:129–142, 2010.

W. Li, Y. Gu, L. Liu, and E. Tchetgen Tchetgen.
Demystify the multiple robust estimators.
In Preparation.

L. Liu and M.G. Hudgens.
Large sample randomization inference with interference.
Journal of the American Statistical Association, 109:288–301, 2014.

L. Liu, M.G. Hudgens, and S. Becker-Dreps.
On inverse probability-weighted estimators in the presence of interference.
Biometrika, 103:829–842, 2016.

L. Liu, W. Miao, B. Sun, J. Robins, and E. Tchetgen Tchetgen.
Instrumental variable estimation of the marginal average effect of treatment on the treated.
Submitted, 2015.

L. Liu and E. Tchetgen Tchetgen.
Indirect adjustment for homophily bias with a negative control variable in peer effect analysis.
In Preparation.

B. Sun, W. Miao, L. Liu, J. Robins, and E. Tchetgen Tchetgen.
Doubly robust instrumental variable estiamtion in missing not at random problems.
Major revision at Statistical Sinica.

Y. Yin, L. Liu, and Z. Geng.
Assessing the treatment effect heterogeneity with a latent variable.
Statistical Sinica, 2017.

Y. Yin, L. Liu, Z. Geng, and P. Luo.
Optimal criteria to exclude the surrogate paradox and sensitivity analysis.
Under Review at JASA, 2017.

Acknowledgments

Yunjian Yin

Thank you all for coming

Feel free to let me know if you encounter any causal problem in your research

Excluding the Paradox

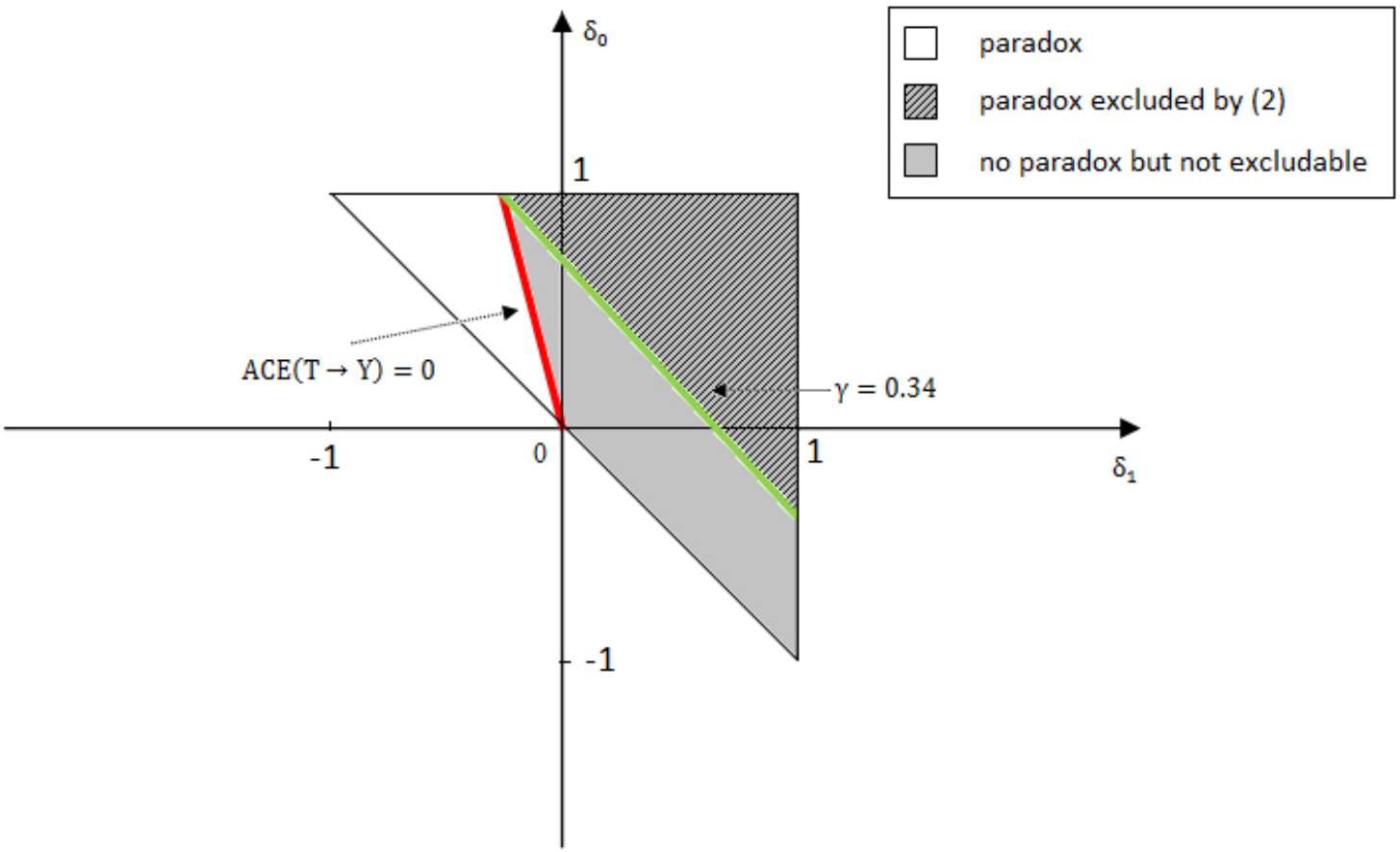


Figure: Partition of the parameter space of (δ_0, δ_1)

Notation

- Binary T treatment, Y primary outcome, S surrogate endpoint.
- U unmeasured confounder that affects both S and Y
- S_t the potential outcome of surrogate if $T = t$
- Y_{ts} the potential outcome if $T = t$ and $S = s$
  - We may also use the notation $Y_{T=t}$ as the potential primary outcome when the intervention is only to set $T = t$
- **Parameter of Interest:**

$$\textbf{ACE}(T \to Y) = P(Y_{T=1} = 1) - P(Y_{T=0} = 1)$$

- **Assumption 1.** (Randomization) $T \perp (Y_{00}, Y_{01}, Y_{10}, Y_{11}, S_0, S_1, U)$

Optimality

Apart from the testability, our criterion also has the following optimality.

- **Definition**
  A criterion to exclude the surrogate paradox is *optimal* if
  - (i) when the criterion is satisfied, the surrogate paradox is absent
  - (ii) when the criterion is not satisfied, one can always find a data generating mechanism that yields the same observed data but suffers from the surrogate paradox.
- That is, one cannot exclude the possibility of surrogate paradox according to the observed data.

Optimality

- Intuitively, an ideal criterion to exclude the surrogate paradox will be based on a sufficient and "almost necessary" condition.
  - The sufficiency gives the condition enough **strength** to rule out surrogate paradox: if the condition is satisfied, the surrogate paradox is guaranteed to be absent.
  - The "almost necessity" gives the condition enough **sharpness** to hold as long as the observed data could rule out the possibility of surrogate paradox: if the condition fails, there exists a data generating process (a set of parameters) with surrogate paradox that can generate the same observed data.

Optimality

- The "almost necessity" differs from necessity in the sense that a necessary condition would require a criteria to rule out the possibility of surrogate paradox whenever it is absent.
- Such necessity is impossible to achieve due to non-identification.
- More specifically, we can only identify a set of data-generating process that is consistent with the observed data.
  - If and only if none of these data generating mechanisms has surrogate paradox, the criterion enable us to exclude surrogate paradox.
- The optimality requires a criterion to capture all the information in the *observed data* to exclude the surrogate paradox.