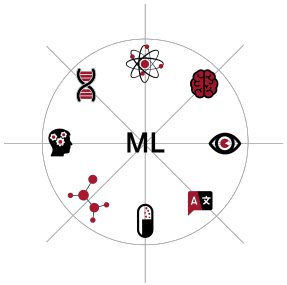
Causal Discovery, Randomization and Individualized Treatment

Konstantin Genin
Group Leader: "Epistemology and Ethics of ML"

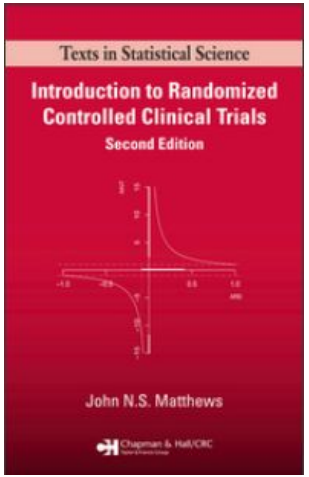
In collaboration with
Conor Mayo-Wilson, University of Washington (Seattle)



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

The Randomized, Controlled Trial (RCT)

“The RCT is the introduction of scientific method into the process of comparing treatments”



The Randomized, Controlled Trial (RCT)

Attempts to discover the relative effectiveness of a new intervention over standard treatment or placebo. Patients are assigned to the different “arms” of the trial by a randomization device.

- Widely considered the “gold standard” research design;
- Typically necessary for FDA approval;
- Raises a number of tricky ethical issues.

A call for RCTs for ML models in Clinical Settings

Perspective

Evaluating artificial intelligence in medicine: phases of clinical research

Yoonyoung Park[1], Gretchen Purcell Jackson[2,3], Morgan A. Foreman[1], Daniel Gruen[1], Jianying Hu[4] and Amar K. Das[1]

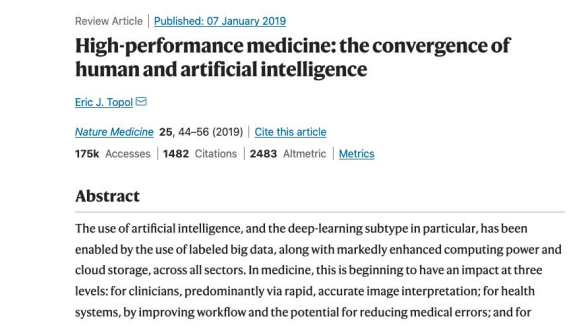
[1]Center for Computational Health, IBM Research Cambridge, Cambridge, Massachusetts, USA, [2]Center for AI, Research, and Evaluation, IBM Watson Health, Cambridge, MA, USA, [3]Department of Pediatric Surgery, Vanderbilt University Medical Center, Nashville, Tennessee, USA and [4]Center for Computational Health, IBM T.J. Watson Research Center, Yorktown Heights, New York, USA

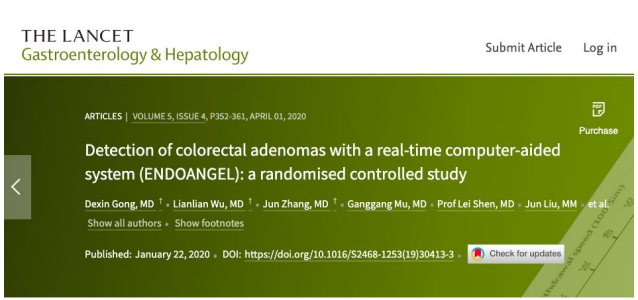
Corresponding Author: Yoonyoung Park, IBM Research, 75 Binney Street, Cambridge, MA 02142, USA (yoonyoung.park@ibm.com)

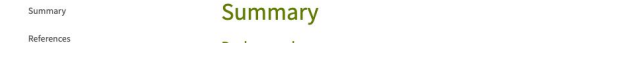
Received 11 September 2019; Revised 15 May 2020; Editorial Decision 25 June 2020; Accepted 1 July 2020

ABSTRACT

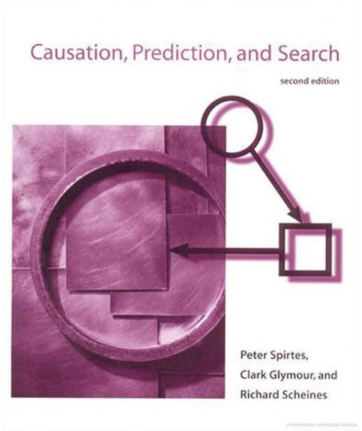
Increased scrutiny of artificial intelligence (AI) applications in healthcare highlights the need for real-world evaluations for effectiveness and unintended consequences. The complexity of healthcare, compounded by the user- and context-dependent nature of AI applications, calls for a multifaceted approach beyond traditional in silico evaluation of AI. We propose an interdisciplinary, phased research framework for evaluation of AI imple-



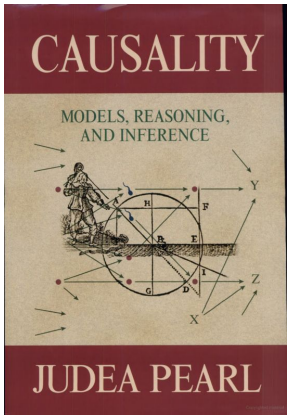




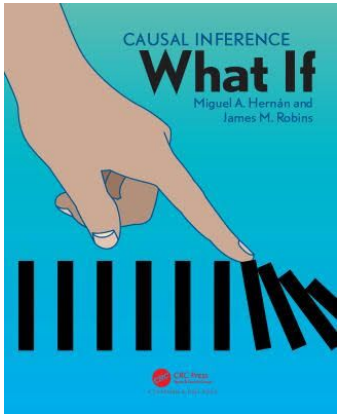
A Proliferation of ML methods for Causal Discovery



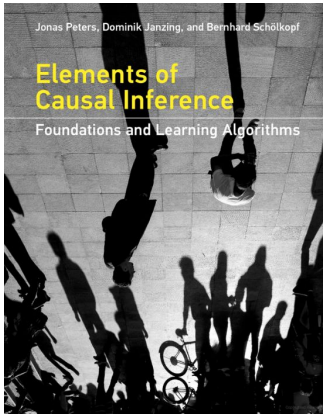
1993/2001



2000



2010



2017

Between Morals and Methodology

If the ethical costs of RCTs are justified, it must be in virtue of the **epistemic superiority** of the randomized, controlled design.

- Exactly what is the epistemic good which RCTs secure?

- Can these goods not be secured with some **other** methodology? Preferably without the same ethical costs?

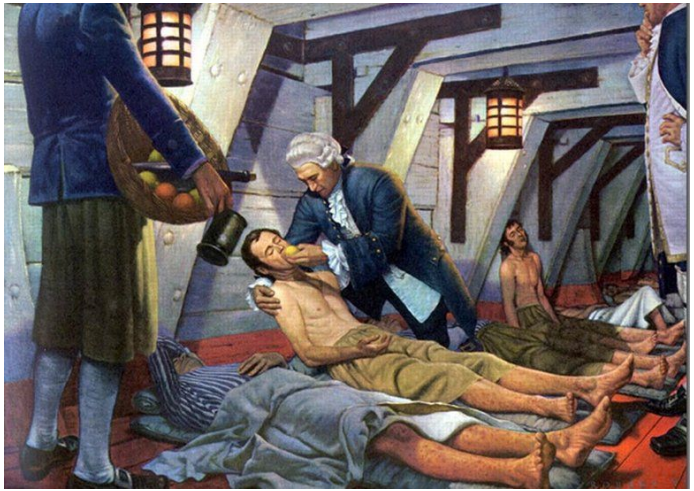
An Early *Controlled* Trial

1747: James Lind, surgeon aboard the HMS Salisbury treats 12 sailors, receiving the same rations, suffering from scurvy.

2 with cider; 2 with seawater; 2 with horseradish;

2 with vinegar; 2 with sulfuric acid

and 2 with **lemons and oranges.**



The First *Randomized* Controlled Trial

1948: A. Bradford Hill, facing a shortage of streptomycin, runs the first randomly allocate trial of streptomycin for tuberculosis.

1965: Bradford Hill proposed a set of nine criteria for epidemiological evidence of a causal relationship. Henceforth widely known as the **Bradford Hill criteria.**

Early Randomizers

1935: Fisher's publishes *The Design of Experiments.*

1925: Fisher publishes *Statistical Methods for Research Workers.*

1921: Fisher publishes *Studies in Crop Variation.*

1883: C.S. Peirce and Jastrow perform randomized experiments in psychophysics.

1780: Charles Deslon proposes a randomized trial to test Mesmer's claims.

The Trouble with Randomization

Randomization comes into prima facie conflict with **therapeutic obligation:**

“A physician should not recommend for a patient therapy such that, given present medical knowledge, the hypothesis that the particular therapy is inferior to some other therapy is more probable than the opposite hypothesis” (Marquis, 1983).

The Trouble with Randomization

Randomization comes into prima facie conflict with **individualized treatment:**

“Although a patient who has been enrolled as a research subject in a RCT may benefit from the therapeutic effects of the treatment being tested, the fact that the treatment cannot be entirely tailored to that patient’s special needs seems to violate the physician’s obligation of unqualified fidelity to his patient’s health” (Schafer, 1983).

Clinical Equipoise

Since theoretical equipoise is very fragile, Freedman proposes **clinical equipoise** instead, which obtains when

“[t]here exists (or, … may soon exit) an honest, professional disagreement among expert clinicians about the preferred treatment” (1987, 144).

The Tragic View of Clinical Research

The discussion around clinical equipoise presupposes

- There is some valuable **epistemic good** secured by randomization;
- **Any** trial methodology which secures this good must **inevitably** come into conflict with the requirements of individual treatment**.**

The Tragic View of Clinical Research

The job of clinical ethics is to reconcile clinicians to this tragic situation:

“These clinical instincts, while understandable and laudable, have the potential to obscure the true nature of clinical research, as investigators and participants alike try to convince themselves that clinical research involves nothing more than the provision of clinical care. One way to try to address this collective and often willful confusion would be to identify a justification for exposing research participants to net risks for the benefit of others.” (Wendler, 2021).

The Tragic View of Clinical Research

But is the tragic view **right?**

- What is the valuable **epistemic good** secured by randomization?

- Is there really **no** methodology that reconciles this good with the ethical requirements of individualized treatment?

Critics of Randomization

Randomization has come in for criticism on purely epistemic grounds.

- Bayesians have a hard time rationally reconstructing randomization (Savage 1961,1962; Kadane & Seidenfeld, 1999; Kasy 2016).

- The theory of optimal design does not endorse randomization (Kiefer 1959; Harville 1975).

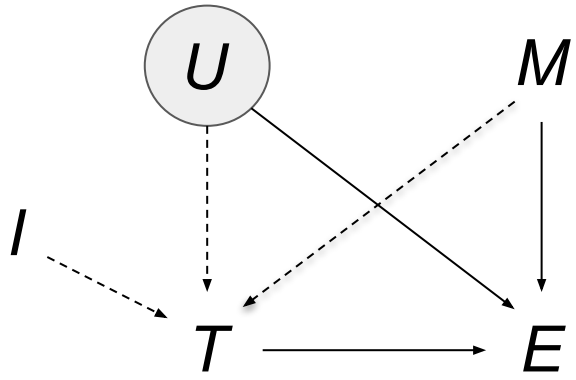
- Philosophers of science have criticized the coherence of randomization (Urbach 1985; Worrall 2002).

Randomization On its Own Terms

What is the best frequentist justification for randomization?

The Causal Situation

T := treatment (binary);
E := effect (binary);
M := measured covariates;
U := unmeasured covariates;
I := randomizer.



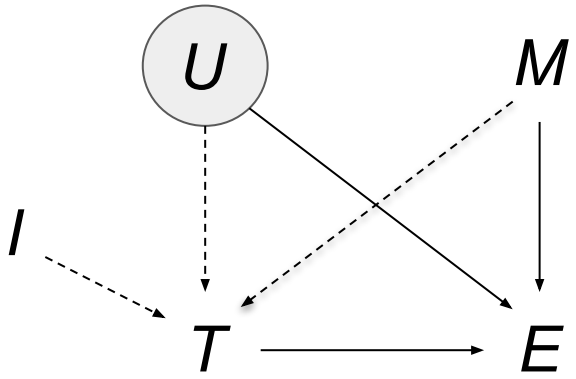
Average Treatment Effect

The goal is to estimate the
average treatment effect (ATE):

$$P(E = 1|do(T = 1)) - P(E = 1|do(T = 0))$$

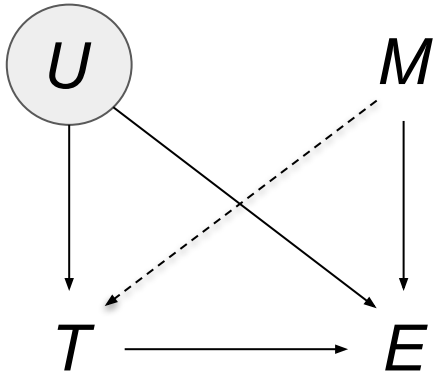
Or, in the notation of the potential outcomes framework:

$$\frac{1}{n}\sum_{i \leq n} P(E_i^{t=1} = 1) - P(E_i^{t=0} = 1)$$



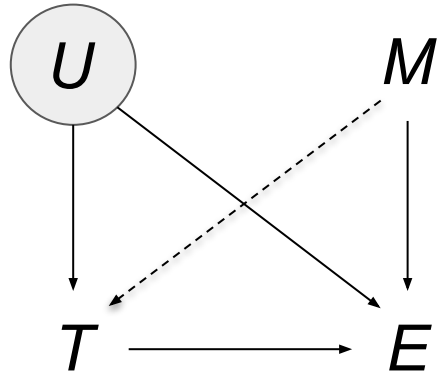
Trouble with Observational Studies

If there is an unobserved common cause of *T, E*
it is easy to come up with examples in which the ATE
is **not identified.**

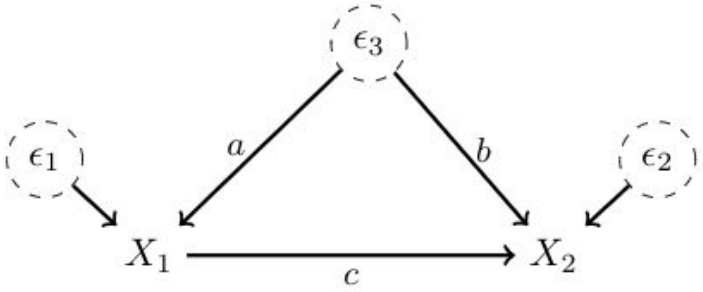


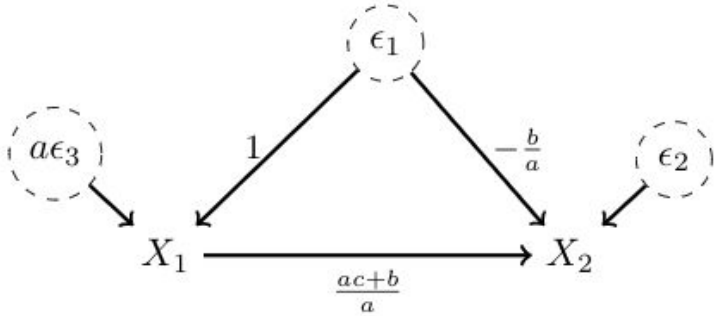
Trouble with Observational Studies

If there is an unobserved common cause of *T, E*
it is easy to come up with examples in which the ATE
is **not identified.**



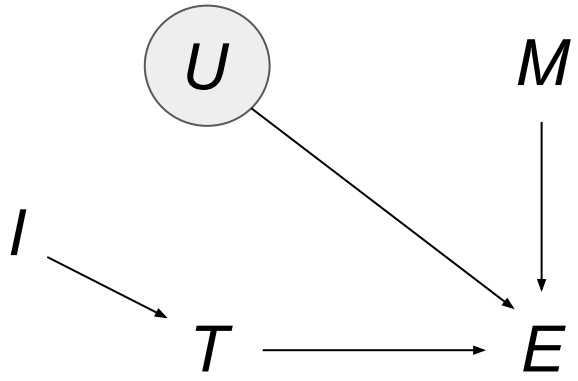
Trouble with Observational Studies





The Point of Randomization

Randomization “breaks edges” into treatment, so that any association between *T* and *E* is due to the causal effect of *T* on *E* and not shared common causes.

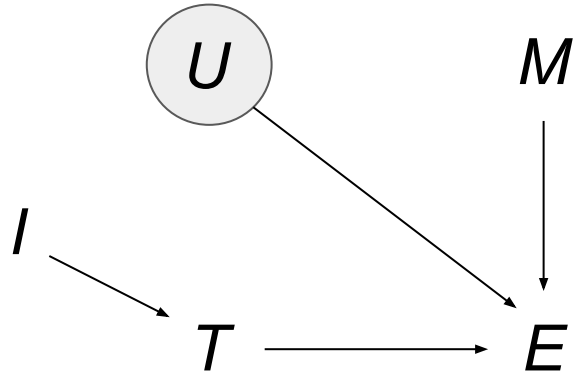


The Point of Randomization

It ensures that the ATE is **identified** and equal to

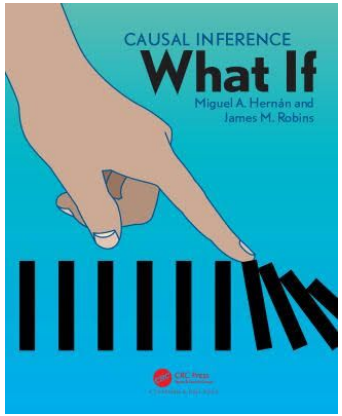
$$P(E = 1|T = 1) - P(E = 1|T = 0)$$

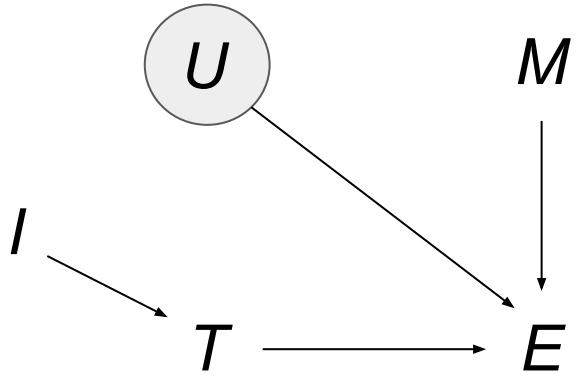
Moreover an **unbiased estimate** of the ATE is easily obtained.



The Point of Randomization

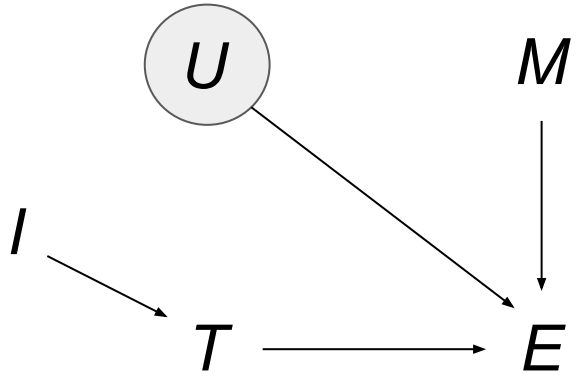
“In ideal randomized experiments, association is causation”





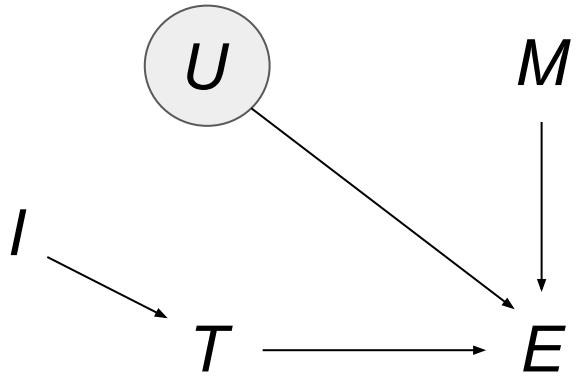
No Other Way?

But is breaking edges into *T* the only way to render the ATE identified and construct unbiased estimates?



No Other Way?

But is breaking edges into *T* the only way to render the ATE identified and construct unbiased estimates?



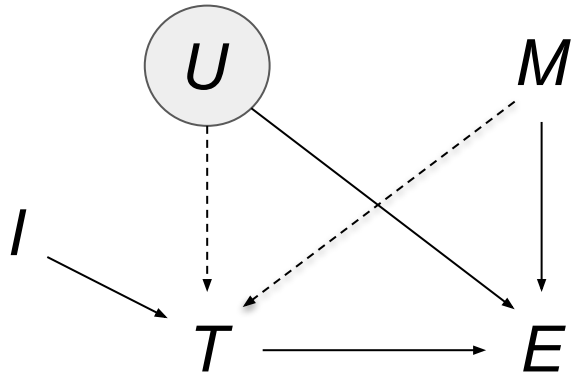
No!

Instrumental Variables

I is an **instrumental variable** if (roughly)

- *I* is statistically independent of *U,M;*
- the only unblocked path from *I* to *E* goes through *T*

  (a path is blocked if it contains a sequence like … → *T* ← …).

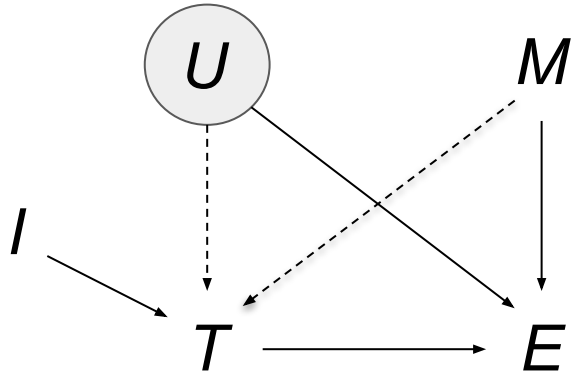


Instrumental Variables

Suppose that

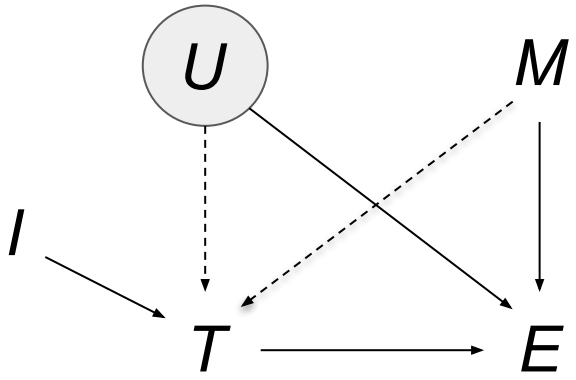
- physicians assign patients to treatment according to their therapeutic judgement
- and only consult a randomizing device (*I*) when they are in equipoise

then *I* is an instrumental variable.



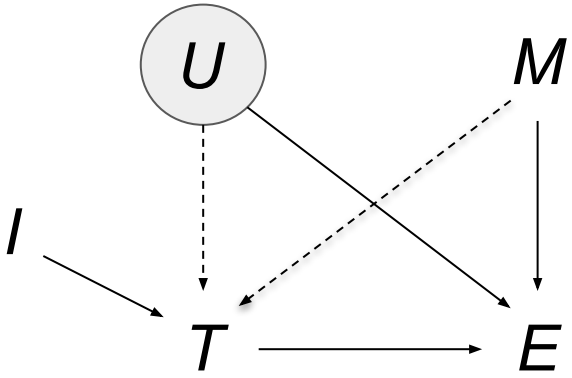
Instrumental Variables

Theorem (Angrist and Imbens 1995): When an instrumental variable satisfies a “monotonicity” condition, then the ATE is **identified** and there is an **unbiased estimator** of the ATE.



Instrumental Variables

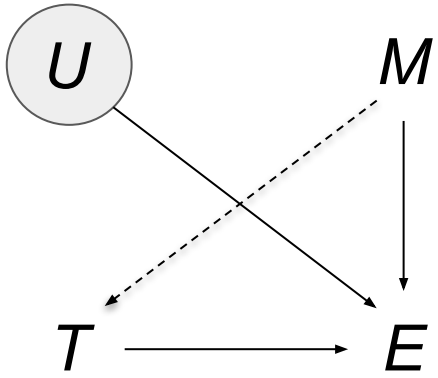
Theorem (Angrist and Imbens 1995): When an instrumental variable satisfies a “monotonicity” condition, then the ATE is **identified** and there is an **unbiased estimator** of the ATE.



Backdoor Adjustment

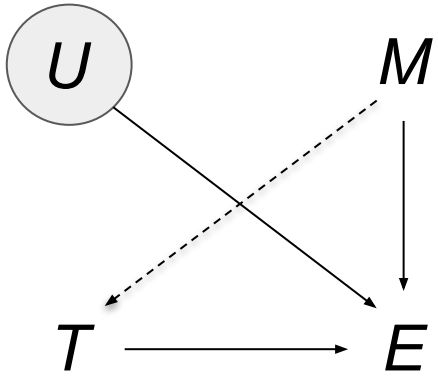
M satisfied the backdoor criterion w.r.t (*T, E)* if

- *M* is not a descendant of *T*;
- *M* blocks every path between *T* and *E* that has an arrow into *T.*

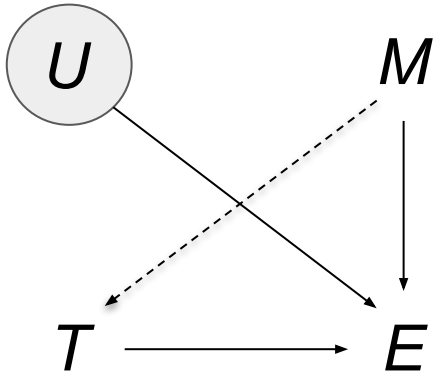


Backdoor Adjustment

Theorem (Pearl, 1993) If there is observed variable *Z* satisfying the backdoor criterion wrt (*T, E)*, then it is possible to construct an unbiased estimate of the causal effect of *T* on *E.*



Backdoor Adjustment

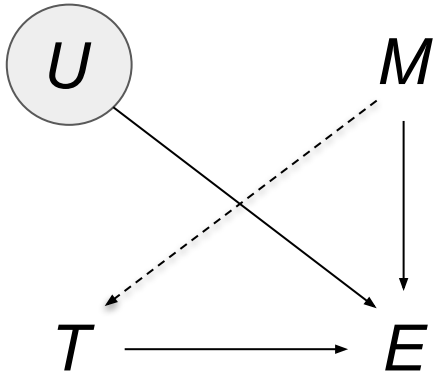


Suppose that

- physicians make assignment to treatment **only** on the basis of observed covariates *M*,

then *M* satisfies the backdoor criterion wrt (*T, E*).

Backdoor Adjustment



Suppose that

- physicians make assignment to treatment **only** on the basis of observed covariates *M*,

then *M* satisfies the backdoor criterion wrt (*T, E*).

Cut-off Designs

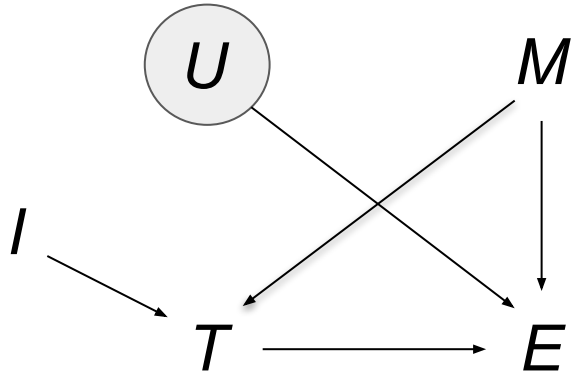
Clinicians rate patients on a continuous scale according to disease severity then

- assign low/high severity patients to less/more aggressive treatment, respectively;

- Randomize patients with moderate severity.

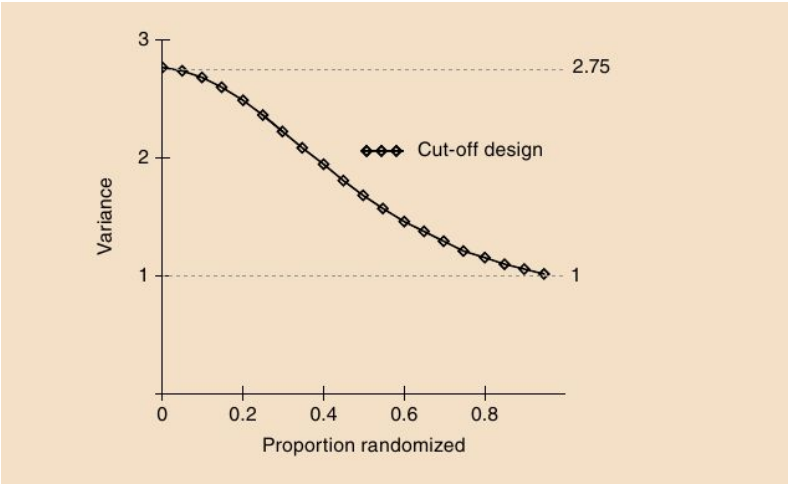
Estimate the ATE by **ridge regression.**

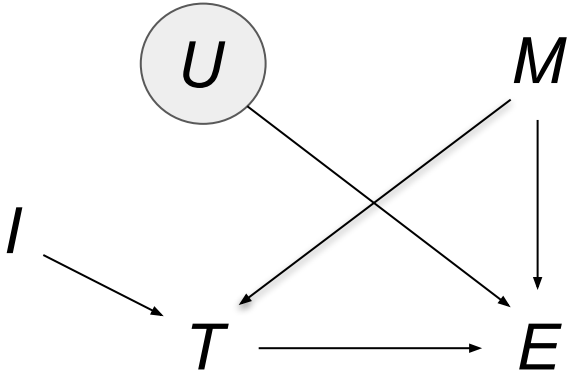
(See Cappelleri, 1995).



Cut-off Designs

Senn (2008) criticizes this design on grounds of **efficiency.**





Randomization On its Own Terms

Neither guaranteeing that

1. the ATE is identified, nor that
2. there exists an unbiased estimator of the ATE,

is sufficient to justify randomization.

Other designs get the same goods and are less hostile to individualized treatment.

Randomization On its Own Terms

If there is a frequentists argument justifying randomization over other methods, it **cannot** be framed in terms of identification or unbiased estimation.

It must be about **efficiency**.

I.e. the **variance** of the estimator.

Randomization On its Own Terms

Are there such arguments?

Minimax Justifications

A series of somewhat neglected papers (Wu 1981; Li 1983; Waite and Woods 2020) develops a **minimax risk** argument for randomization.

Minimax Justifications

Suppose that for each patient, the effect of treatment is given by:

$$E_{it} = \alpha_t + g_i + \epsilon_{ut}$$

effect of treatment *t* on patient *i*

fixed effect of treatment *t*

patient effect

Independent mean-zero noise

Minimax Justifications

Let **g** = (g_1, g_2, …, g_N) be an assignment of patient effects to individuals.

Let **G** be the set of all assignments consistent with background knowledge.

Symmetry assumption: if **g** is in **G**, then so is every permutation of **g**.

Minimax Justifications

Theorem (Wu, 1981) The fully randomized design minimizes the maximum MSE of the estimate of the α_t over all possible values of **G**.

So the fully randomized design has the best worst-case efficiency.

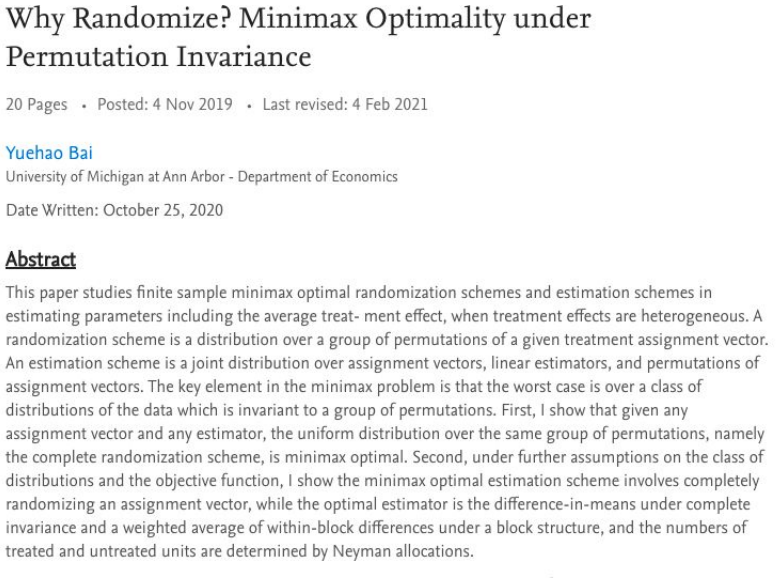
Research Questions

Can these minimax arguments be generalized away from the linearity assumptions?

Generalized Minimax Justifications

Can these minimax arguments be generalized away from the linearity assumptions?

Yes, to some extent.



Why Randomize? Minimax Optimality under Permutation Invariance

20 Pages • Posted: 4 Nov 2019 • Last revised: 4 Feb 2021

Yuehao Bai
University of Michigan at Ann Arbor - Department of Economics

Date Written: October 25, 2020

Abstract

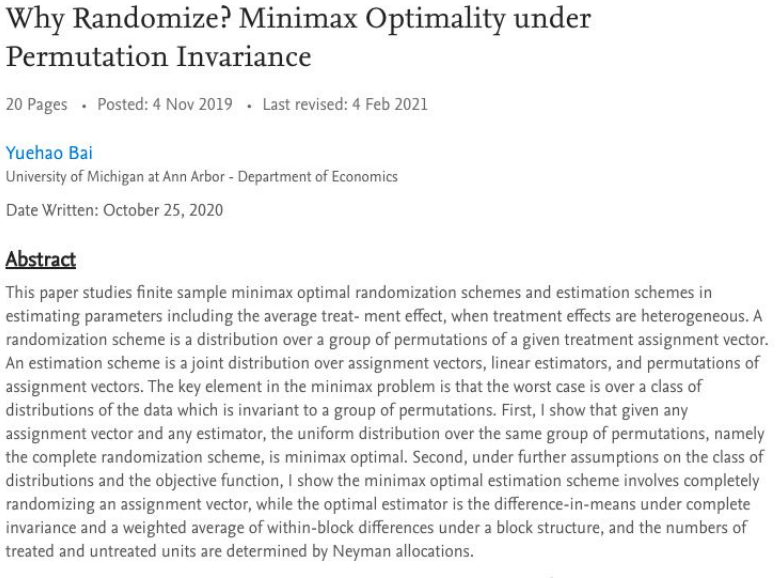
This paper studies finite sample minimax optimal randomization schemes and estimation schemes in estimating parameters including the average treat- ment effect, when treatment effects are heterogeneous. A randomization scheme is a distribution over a group of permutations of a given treatment assignment vector. An estimation scheme is a joint distribution over assignment vectors, linear estimators, and permutations of assignment vectors. The key element in the minimax problem is that the worst case is over a class of distributions of the data which is invariant to a group of permutations. First, I show that given any assignment vector and any estimator, the uniform distribution over the same group of permutations, namely the complete randomization scheme, is minimax optimal. Second, under further assumptions on the class of distributions and the objective function, I show the minimax optimal estimation scheme involves completely randomizing an assignment vector, while the optimal estimator is the difference-in-means under complete invariance and a weighted average of within-block differences under a block structure, and the numbers of treated and untreated units are determined by Neyman allocations.

Generalized Minimax Justifications

Can these minimax arguments be generalized away from the linearity assumptions?

Yes, to some extent.

But the nature of this justification is very different from that suggested by the tragic view!



Why Randomize? Minimax Optimality under Permutation Invariance

20 Pages • Posted: 4 Nov 2019 • Last revised: 4 Feb 2021

Yuehao Bai
University of Michigan at Ann Arbor - Department of Economics

Date Written: October 25, 2020

Abstract

This paper studies finite sample minimax optimal randomization schemes and estimation schemes in estimating parameters including the average treat- ment effect, when treatment effects are heterogeneous. A randomization scheme is a distribution over a group of permutations of a given treatment assignment vector. An estimation scheme is a joint distribution over assignment vectors, linear estimators, and permutations of assignment vectors. The key element in the minimax problem is that the worst case is over a class of distributions of the data which is invariant to a group of permutations. First, I show that given any assignment vector and any estimator, the uniform distribution over the same group of permutations, namely the complete randomization scheme, is minimax optimal. Second, under further assumptions on the class of distributions and the objective function, I show the minimax optimal estimation scheme involves completely randomizing an assignment vector, while the optimal estimator is the difference-in-means under complete invariance and a weighted average of within-block differences under a block structure, and the numbers of treated and untreated units are determined by Neyman allocations.

Research Questions

What is the **precise tradeoff** between individualized treatment and worst-case efficiency of the estimator?

e.g. If we want X% of patients to get individualized treatment, how many more trial participants (*N'*) would we need to achieve the same efficiency as a fully randomized RCT with *N*?

What is more important: giving (most) participants individualized treatment, or getting informative results with fewer participants?

Research Questions

What is the **precise tradeoff** between individualized treatment and worst-case efficiency of the estimator?

If we want X% of patients to get individualized treatment, how many more trial participants (*N'*) would we need to recruit in order to achieve the same efficiency as a fully randomized RCT with *N* patients?

Research Questions

What is the **precise tradeoff** between individualized treatment and worst-case efficiency of the estimator?

What is more important: giving (most) participants individualized treatment, or getting informative results with fewer participants?

Takeaway

If individualized treatment and estimation efficiency trade off, we should be able to say something quantitative about the nature of the trade-off.

The existence of **some** trade-off does not justify abandoning all therapeutic obligations.