Title: Some applications of Causal Inference in the real world

Speakers: Ciaran Gilligan-Lee

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 21, 2023 - 10:00 AM

URL: https://pirsa.org/23040125

Abstract: Causal reasoning is vital for effective reasoning in many domains, from healthcare to economics. In medical diagnosis, for example, a doctor aims to explain a patient's symptoms by determining the diseases causing them. This is because causal relations, unlike correlations, allow one to reason about the consequences of possible treatments and to answer counterfactual queries. In this talk I will present two recent causal inference projects done with my collaborators deriving new algorithms to solve problems that arise when applying causal inference in the real world.

Some applications of Causal Inference in the real world

Ciarán Gilligan-Lee

Spotify & University College London


Dr Anthony Hinton
@TonyHinton2016
Thanks to @abirballan
orange line= vaccinated 10-59 y old
blue line= unvaccinated 10-59 y old
Orange line is ABOVE blue line
=> MORE deaths in vaccinated than unvaccinated?
Is the right policy to vaccinate low-risk populations?
ANY explanation @CMO_England @sajidjavid @ONS ???
Death Rates 10-59 by Vaccine Status ONS Data
Published November 2021
3.5
3
2.5
2
1.5
1
0.5
0
19/03/2021
19/04/2021
19/05/2021
19/06/2021
19/07/2021
19/08/2021
19/09/2021
Death rates per 100,000 in the age range 10-59 Unvaccinated
Death rates per 100,000 in the age range 10-59 After 2nd vaccine
08:58 · 20/11/2021 · Twitter for iPad
1,109 Retweets 148 Quote Tweets 1,706 Likes

Motivating example

- The findings were very weird indeed, flying in the face of medical knowledge and confounding experts

- Yet the finding was irrefutable: death rates for vaccinated people are higher than for unvaccinated people

Do vaccines work?

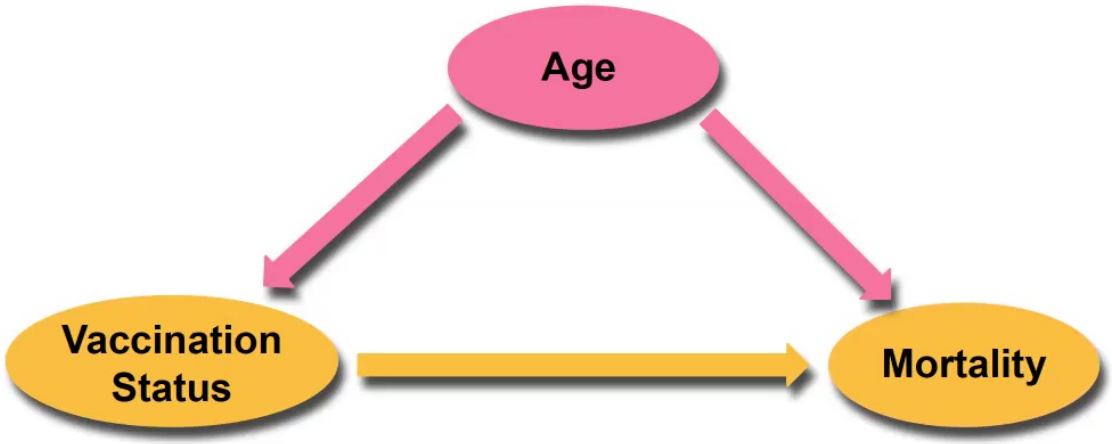
Let's look at a picture

The average unvaccinated person is much younger than the average vaccinated person.

Therefore they have a lower mortality rate. Any benefit from the vaccines is swamped by the increase in mortality with age!

Age is a **confounder** between Vaccine Status and Mortality

When we control for age, vaccinations are shown to reduce mortality rate.



Motivating example

- Any action or policy change based on these correlations—such as whether to vaccinate—would not increase patient survival.

Take home: Relying on correlations extracted from observational data can lead to embarrassing, costly, and dangerous mistakes.

- To overcome this, we need to understand cause and effect

Why is this important for Spotify?

Why is this important for Spotify?

- Usually randomised controlled trials or A/B tests tell us about cause & effect.

- But sometimes A/B tests can't be performed. They could be too damaging to user experience, or technically too hard to implement:
  - "Do app crashes cause churn?"
  - "Does podcast or audiobook consumption cause retention?"

- **Causal Inference** provides a set of methods and tools for learning and quantifying cause and effect, even without A/B tests – given some assumptions.

Disentangling joint-interventions

- In many applications, only a single intervention is possible at a given time, or interventions are applied one after another in a sequential manner
- However, in some areas, **multiple interventions are concurrently applied**:
  - in medicine, patients that possess many commodities may have to be simultaneously treated with multiple prescriptions;
  - in computational advertising, people may be targeted by multiple concurrent campaigns, and so on.
  - during the pandemic, many interventions were applied at same time, e.g. mask wearing, work from home, schools closed, etc.
- How can we learn the individual effect of each intervention?

Problem: disentangling interventions

Given samples from observational and joint-interventions data

$$\mathbb{E}[Y|X_i = x_i, X_j = x_j, C = c], \text{ and } \mathbb{E}[Y|\text{do}(X_i = x_i, X_j = x_j), C = c]$$

When can we learn, or **identify**, conditional average causal effects of single-interventions

$$\mathbb{E}[Y|\text{do}(X_i = x_i), X_j = x_j, C = c], \text{ or } \mathbb{E}[Y|X_i = x_i, \text{do}(X_j = x_j), C = c]$$

Identifiability

A quantity is identifiable from a specific type of data if every model that agrees on that data produces the same value for the quantity

Hence, if two models agree on the data, but not on the quantity, then it is not identifiable from that data

This is not identifiable in general

Intuition:

- All variables are binary, and all latents are perfectly correlated.
- Observationally, the models look the exact same! Moreover as Y is the same function of X's in both models, joint-interventions are the same
- But when we intervene on X_1, X_2 behaves differently in both models—as X_2 doesn't causally depend on X_1 in M', but it does in M.
- Hence observations and joint-interventions are not enough to fully constrain single-interventions.

That is, we need more assumptions for identifiability

$\mathcal{M}$	$\mathcal{M}'$
$X_1 = U_1$	$X_1 = U_1$
$X_2 = X_1 U_2$	$X_2 = U_2$
$Y = X_1 X_2 U_Y$	$Y = X_1 X_2 U_Y$

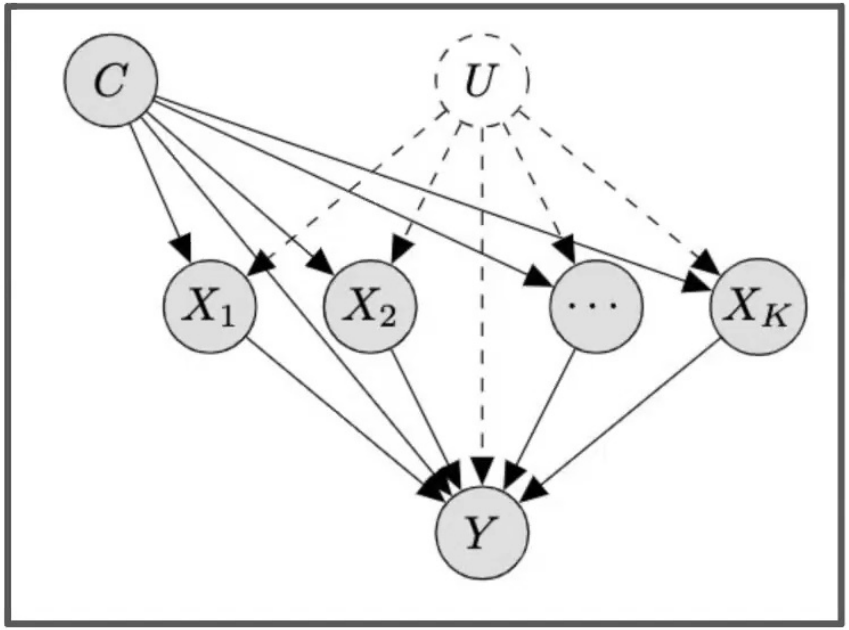
U_1 = U_2 = U_y
perfectly correlated bits

But it ***is*** identifiable from ***extra assumptions***

Theorem 2 (Identifiability of disentangled conditional average treatment effects in additive noise models with symmetric structure)**.**
Let $\mathcal{M} = \langle \{\boldsymbol{C}, \boldsymbol{X}, Y\}, \boldsymbol{U}, \boldsymbol{f}, \mathsf{P}_U \rangle$ *be an SCM, where*

$$X_i = f_i(\boldsymbol{C}) + U_i, \quad \forall i = 1, \ldots, K,$$
$$Y = f_Y(\boldsymbol{C}, \boldsymbol{X}) + U_Y,$$

$C \perp\!\!\!\perp U$, *and* $\mathsf{P}_U \sim \mathcal{N}(0, \Sigma)$. *The estimand* $\mathbb{E}[Y|do(X_i), C]$ *is identifiable from the conjunction of two data regimes: (1) the observational distribution, and (2) any interventional distribution on a set of treatments* $\boldsymbol{X}_{int} \subseteq \boldsymbol{X}$ *that holds* X_i*:* $X_i \in \boldsymbol{X}_{int}$.

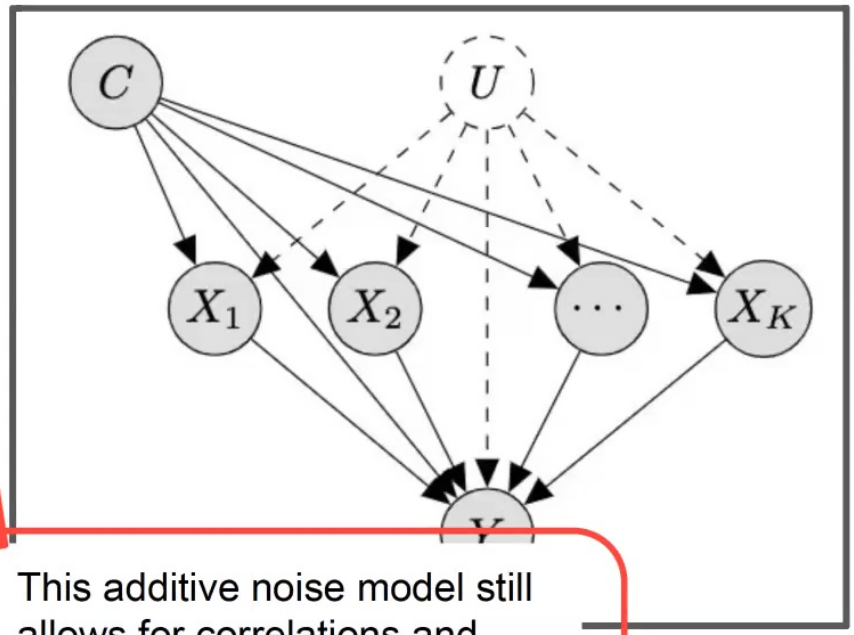


But it ***is*** identifiable from ***extra assumptions***

Theorem 2 (Identifiability of disentangled conditional average treatment effects in additive noise models with symmetric structure)**.**
Let $\mathcal{M} = \langle \{\boldsymbol{C}, \boldsymbol{X}, Y\}, \boldsymbol{U}, \boldsymbol{f}, \mathsf{P}_U \rangle$ *be an SCM, where*

$$X_i = f_i(\boldsymbol{C}) + U_i, \quad \forall i = 1, \ldots, K,$$
$$Y = f_Y(\boldsymbol{C}, \boldsymbol{X}) + U_Y,$$

$C \perp\!\!\!\perp U$, *and* $\mathsf{P}_U \sim \mathcal{N}(0, \Sigma)$. *The e[illegible]and* $\mathbb{E}[Y|do(X_i), C]$ *is identifiable from the conjunctio[illegible]f two data regimes: (1) the observational distribution, an[illegible] (2) any interventional distribution on a set of treatments* $\boldsymbol{X}_{int} \subseteq \boldsymbol{X}$ *that holds* X_i*:* $X_i \in \boldsymbol{X}_{int}$.



This additive noise model still allows for correlations and interactions between treatments, through observed and unobserved confounders

Learning algorithm based on results

Estimating a structural causal model from a combination of observational and interventional data boils down to:

1. estimating the structural equations,
2. estimating the noise distribution

$$\mathbb{E}[Y|\boldsymbol{C}; \mathrm{do}(\boldsymbol{X}_{\mathrm{int}}); \boldsymbol{X}_{\mathrm{obs}}] = f_Y(\boldsymbol{C}; \boldsymbol{X}) + \mathbb{E}[U_Y|\boldsymbol{X}_{\mathrm{obs}}].$$

We employ an Expectation-Maximisation-style iterative algorithm to achieve this

Full details in the paper…

$$L(x_i; \theta, \Sigma) = \mathsf{P}_U(x_i - f_i(\mathrm{PA}(x_i); \theta); \Sigma)$$

Algorithm 1 SCM Estimation for Symmetric ANMs

Input: Dataset $\mathcal{D}$
Output: Parameter estimates $\widehat{\theta}$, $\widehat{\Sigma}$
1: Initialise $\widehat{\theta}$ and $\widehat{\Sigma}$
2: **while** not converged **do**
3: // Solve for θ with fixed $\widehat{\Sigma}$
4: Optimise log-likelihood in Eq. 7
5: // Solve for Σ with fixed $\widehat{\theta}$
6: Estimate $\widehat{\Sigma}$ from $\widehat{U} = \boldsymbol{x} - \boldsymbol{f}(\boldsymbol{x}; \widehat{\theta})$
7: **return** $\widehat{\theta}, \widehat{\Sigma}$

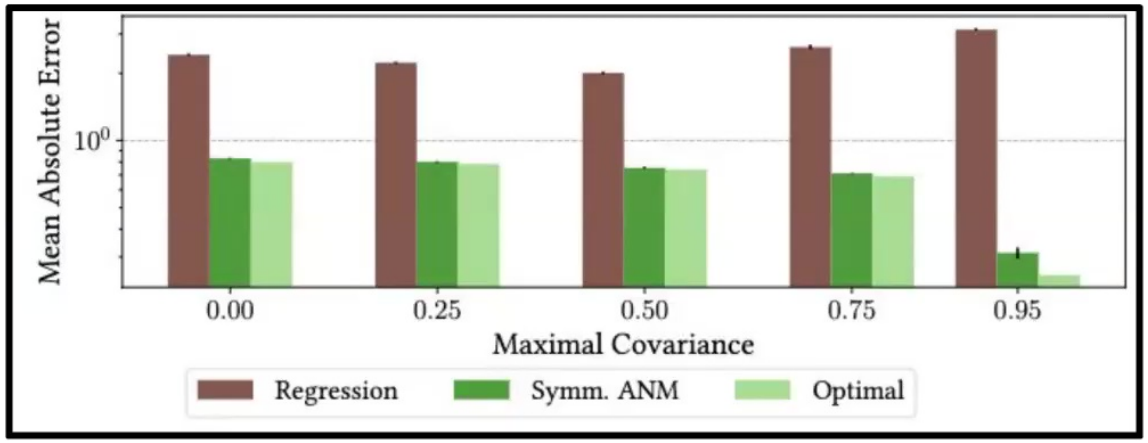
Experiments: How robust is learning to increasing confounding strength?

We showed our method is robust under varying levels of unobserved confounding by testing it in a semi-synthetic setup.

This was based on real-world data from the International Stroke Trial database: a large, randomised trial of up to 14 days of antithrombotic therapy after stroke onset.

There are two possible treatments: aspirin allocation dosage, heparin allocation dosage.

The goal is to understand the effects of these treatments on a composite outcome, a continuous value in [0,1] predicting the likelihood of patients' recovery.



Accepted at NeurIPS 2022 and on arXiv at 2210.05446

Disentangling causal effects from sets of interventions in the presence of unobserved confounders

Olivier Jeunen[2] Ciarán M. Gilligan-Lee Rishabh Mehrotra Mounia Lalmas
Spotify, London, UK

Abstract

The ability to answer causal questions is crucial in many domains, as causal inference allows one to understand the impact of interventions. In many applications, only a single intervention is possible at a given time. However, in some important areas, multiple interventions are concurrently applied. Disentangling the effects of single interventions from jointly applied interventions is a challenging task—especially as simultaneously applied interventions can interact. This problem is made harder still by unobserved confounders, which influence both treatments and outcome. We address this challenge by aiming to learn the effect of a single-intervention from both observational data and sets of interventions. We prove that this is not generally possible, but provide identification proofs demonstrating that it can be achieved in certain classes of additive noise models—even in the presence of unobserved confounders. Importantly, we show how to incorporate observed covariates and learn heterogeneous treatment effects conditioned on them for single-interventions.

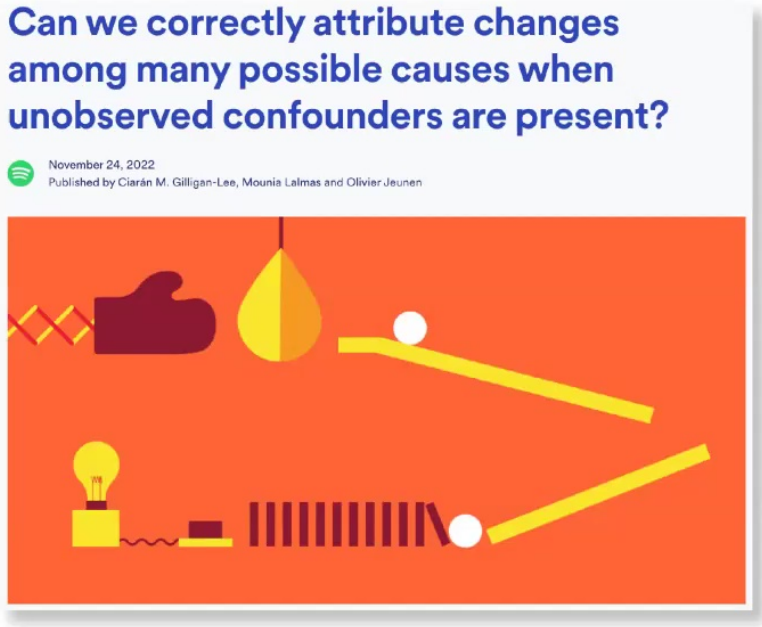
manner. However, in some important areas, multiple interventions are concurrently applied. For instance, in medicine, patients that possess many commodities may have to be simultaneously treated with multiple prescriptions; in computational advertising, people may be targeted by multiple concurrent campaigns; and in dietetics, the nutritional content of meals can be considered a joint intervention from which we wish to learn the effects of individual nutritional components.

Disentangling the effects of single interventions from jointly applied interventions is a challenging task—especially as simultaneously applied interventions can interact, leading to consequences not seen when considering single interventions separately. This problem is made harder still by the possible presence of unobserved confounders, which influence both treatments and outcome. This paper addresses this challenge, by aiming to learn the effect of a single-intervention from both observational data and sets of interventions. We prove that this is not generally possible, but provide identification proofs demonstrating it can be achieved in certain classes of non-linear causal models with additive Gaussian noise—even in the presence of unobserved confounders. Importantly, we show how to incorporate observed covariates, which can be high-dimensional, by learning heterogeneous treatment effects conditioned on them for single-interventions.

Our main contributions are:

1 INTRODUCTION

1. A proof that without restrictions on the causal



Overview on Spotify Research blog:
https://research.atspotify.com/blog/

Examples of disentangling problem at Spotify

- There are a range of playlists/albums/podcasts that are recommended to a user at a given time, what's the individual impact of each one?

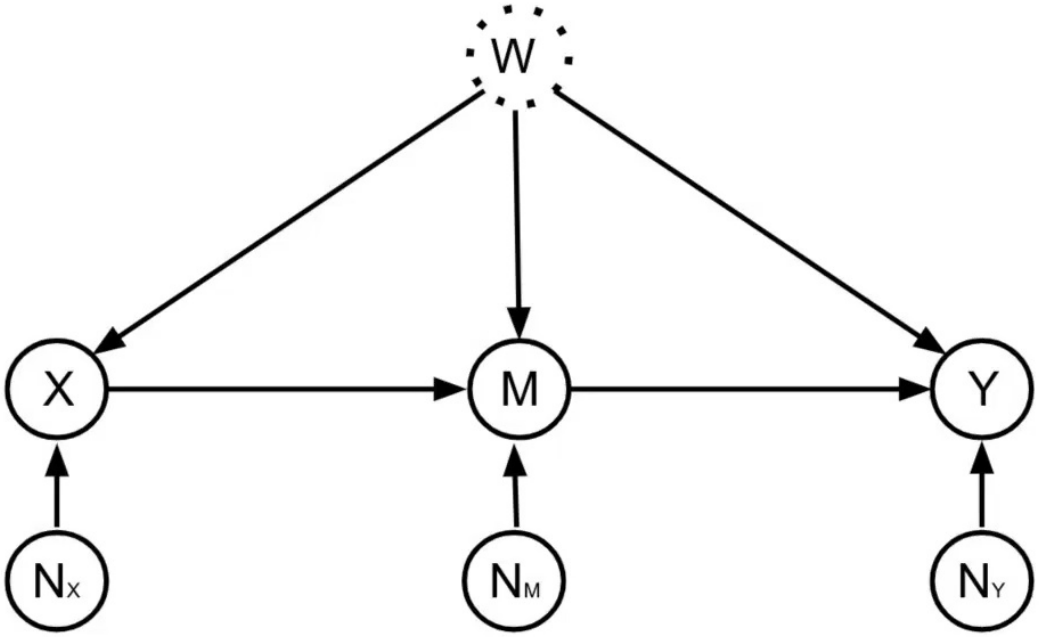
- There are a collection of actions an artist can take to build their fanbase and improve their career, which ones have the biggest effect for a given artist?

And many, many more….

Estimating long-term outcomes

- Even when we can use A/B tests, they usually have relatively short durations due to cost considerations. This makes learning long-term causal effects a very challenging task in practice.
- Often short-term outcomes are different to long-term ones, and, as many decision-makers are interested in long-term outcomes, this is a crucial problem to address.
- For instance:
  - Technology companies are interested in understanding the impact of deploying a new feature on long-term retention
  - Economists are interested in long-term outcomes of job training programs
  - Doctors are interested in the long-term impacts of medical interventions, such as treatments for stroke

The problem



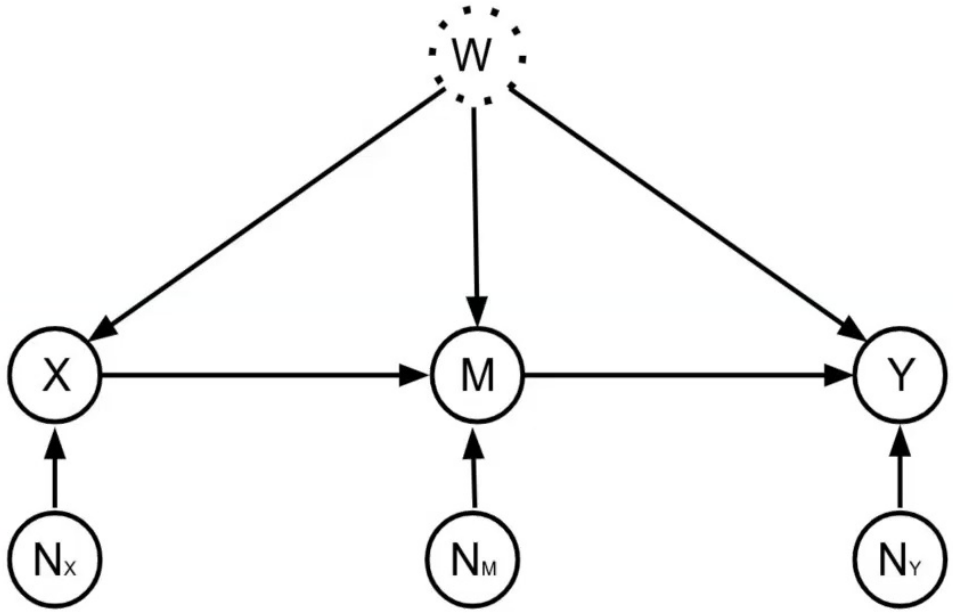
Given experimental samples between (X, M), and (historical) observational samples between (X, M, Y), can we estimate the causal effect of X on Y?

The problem: a concrete example

Suppose X is dosage of a specific drug, M is the severity of symptom after 2 weeks, and Y is the symptom severity after 6 months.

If we have historical observational samples involving X, M, and Y, and we're given samples from a recent experiment between dosage and symptoms two weeks later, can we combine this with the observational samples to estimate the effect of dosage on symptom severity 6 months later?

The issue is that observational samples can be confounded, while experimental samples are not. Moreover, because they're samples from different distributions, we don't observe the long term outcomes for those units in the experiment.



The problem

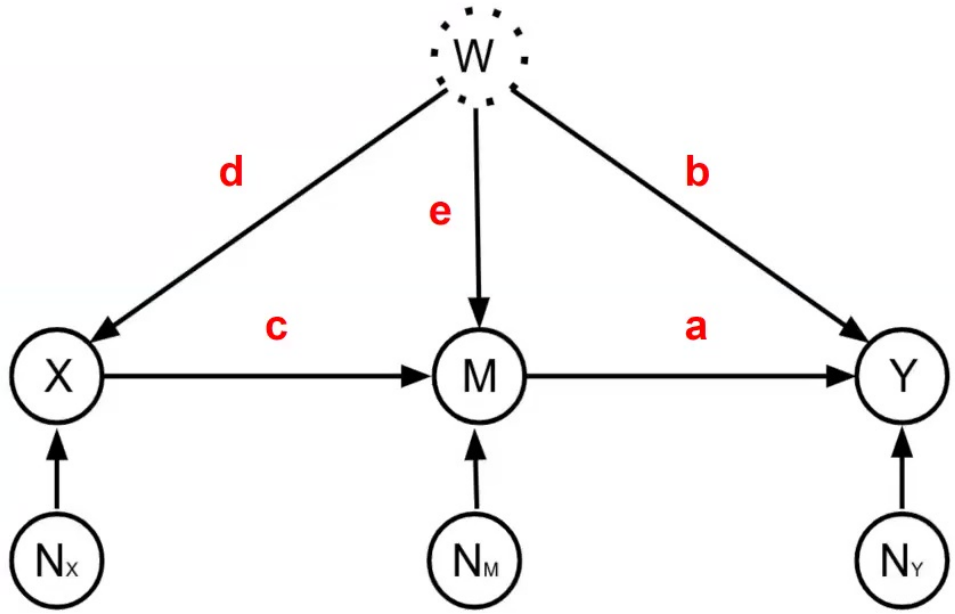
Let's assume a linear Gaussian model:

$$W = N_W$$

$$X = d.W + N_X$$

$$M = c.X + e.W + N_M$$

$$Y = a.X + b.W + N_Y$$



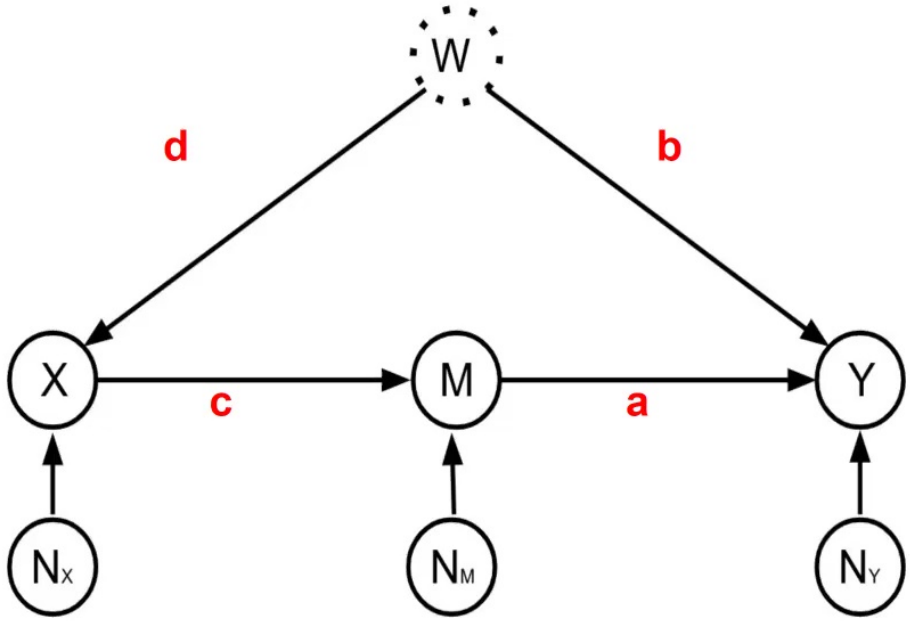
Warm up: front-door structure

Consider the front-door causal structure. Here, to estimate the causal effect of X on Y, from observational data we:

1. Regress M on X to get **c**,
2. Regress Y on M and X to get **a.**

The causal effect is just their product:

a.c

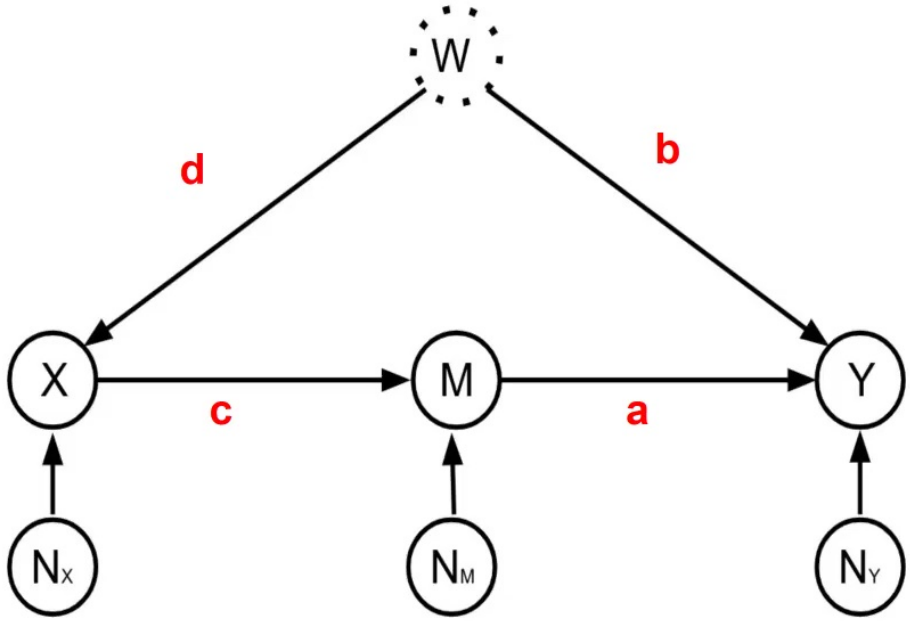


Warm up: new front-door estimator

Instead of the standard estimation strategy from the previous slide, let's try something new. Estimating **c** as before, estimate **a** as follows:

1. Regress M on X, and compute the residual: N_M
2. Use N_M as an instrumental variable for M -> Y

Regress Y on N_M and M on N_M and take the ratio of the coefficients to get **a**.



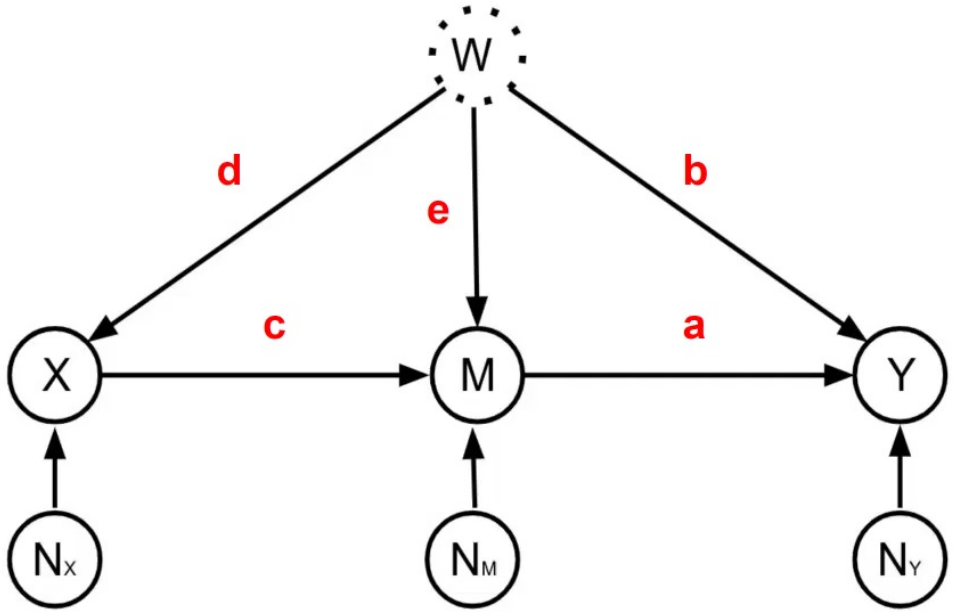
The full problem

The residual from regressing M on X isn't N_M in this case, due to confounding from W.

However, using the experimental samples between X and M, we can remove the confounding bias on the residual, and use this de-biased residual as an instrument for M -> Y.

Instrument =
Residual[M|X] - [E(M-**c**.X)/E(X)].X

This requires obtaining **c** from the experimental dataset

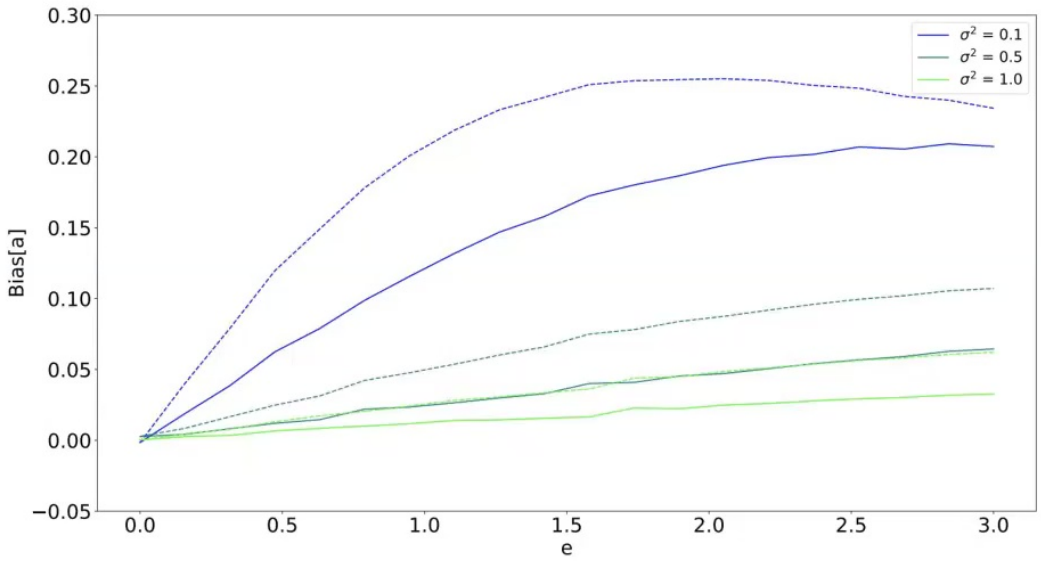


Experiments: How robust is learning to hidden confounding?

We showed our method is robust to unobserved confounding by testing it in a semi-synthetic setup.

This was based on real-world data from the International Stroke Trial database: a large, randomised trial of up to 14 days of antithrombotic therapy after stroke onset.

Using systolic blood pressure at randomisation as treatment, & age as hidden confounder, the goal is to estimate the effect of a synthetic mediator on synthetic outcome: a value in [0,1] —the likelihood of patients' recovery. Solid line our method, dashed line just controls for observed confounders.



Accepted at CLeaR 2023 and on arXiv at 2302.10625

Proceedings of Machine Learning Research vol TBD:1–22, 2023 2nd Conference on Causal Learning and Reasoning

Estimating long-term causal effects from short-term experiments and long-term observational data with unobserved confounding

Graham Van Goffrier UCAPGWG@UCL.AC.UK
Department of Physics and Astronomy, University College London

Lucas Maystre LUCASM@SPOTIFY.COM
Spotify

Ciarán Gilligan-Lee CIARANL@SPOTIFY.COM, CIARAN.LEE@UCL.AC.UK
Spotify & Department of Physics and Astronomy, University College London

Editors: Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

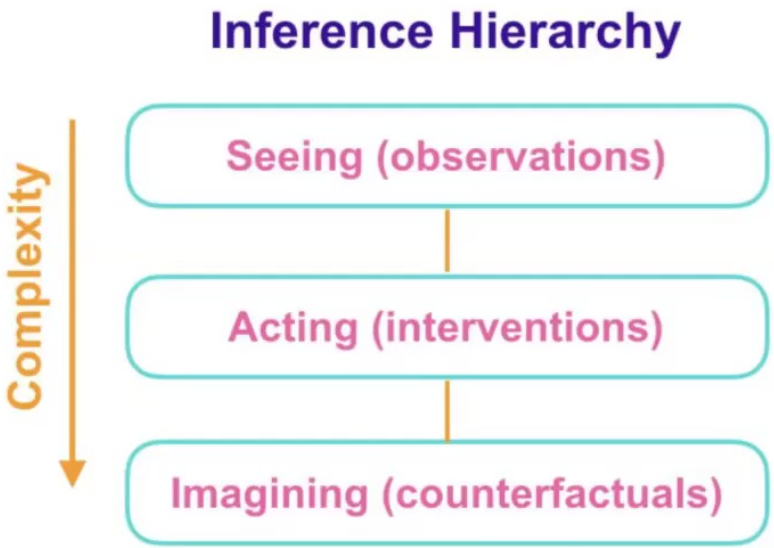
Abstract

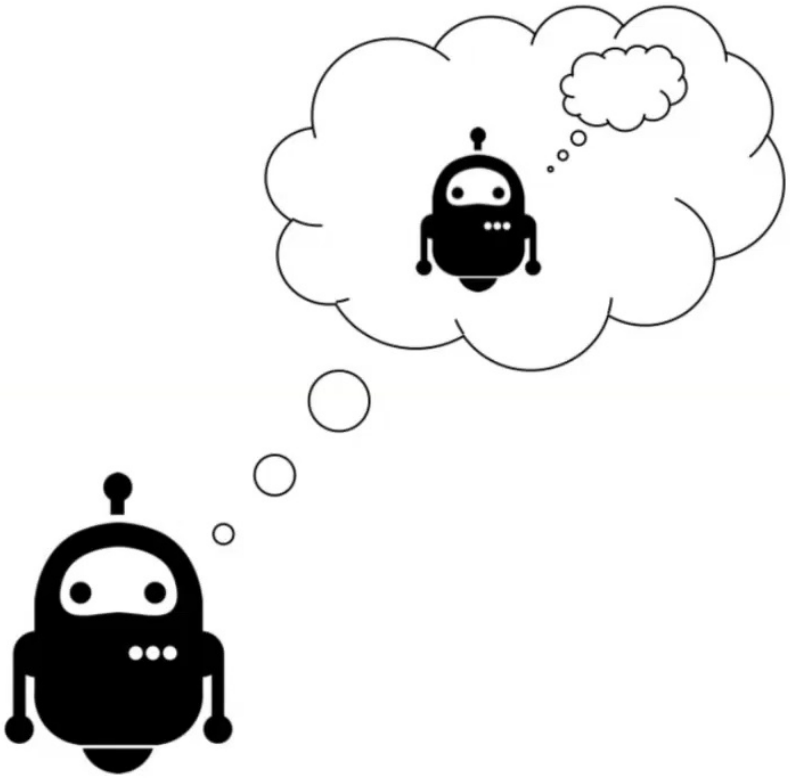
Understanding and quantifying cause and effect is an important problem in many domains. The generally-agreed solution to this problem is to perform a randomised controlled trial. However, even when randomised controlled trials can be performed, they usually have relatively short duration's due to cost considerations. This makes learning long-term causal effects a very challenging task in practice, since the long-term outcome is only observed after a long delay. In this paper, we study the identification and estimation of long-term treatment effects when both experimental and observational data are available. Previous work provided an estimation strategy to determine long-term causal effects from such data regimes. However, this strategy only works if one assumes there are no unobserved confounders in the observational data. In this paper, we specifically address the challenging case where unmeasured confounders are present in the observational data. Our long-term causal effect estimator is obtained by combining regression residuals with short-term experimental outcomes in a specific manner to create an instrumental variable, which is then used to quantify the long-term causal effect through instrumental variable regression. We prove this estimator is unbiased, and analytically study its variance. In the context of the front-door causal structure, this provides a new causal estimator, which may be of independent interest. Finally, we empirically test our approach on synthetic-data, as well as real-data from the International Stroke Trial.

Keywords: Long-term causal effects, latent confounding, linear Structural Causal Models

Overview on Spotify Research blog:
https://research.atspotify.com/blog/

We've only discussed interventional questions…



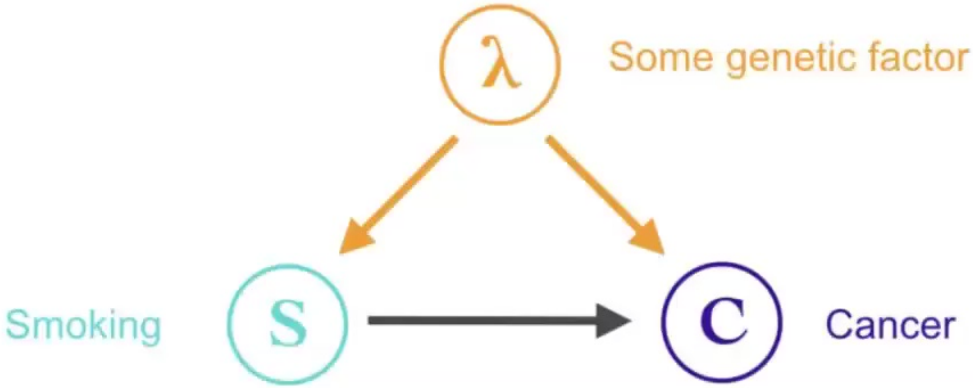


Are counterfactuals useful for anything in practice?

Seeing (observations)

Acting (interventions)

Imagining (counterfactuals)



given

Probability of subject not having cancer

subject has cancer

and if subject was made to not smoke

$$P(C = F | C = T, \mathrm{do}(S = F))$$

"Given subject has cancer, what is the chance they wouldn't if they didn't smoke?"

Yes! Formulating medical diagnosis as a counterfactual task results in expert clinical accuracy

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › articles › article

Article | Open Access | Published: 11 August 2020

Improving the accuracy of medical diagnosis with causal machine learning

Jonathan G. Richens ✉, Ciarán M. Lee & Saurabh Johri

Nature Communications **11**, Article number: 3923 (2020) | Cite this article

72k Accesses | **105** Citations | **952** Altmetric | Metrics

NEWSLETTERS
Sign up to read our regular email newsletters

NewScientist

News Podcasts Video Technology Space Physics Health More Shop Courses Events

AI mimics the way doctors think to make better medical diagnoses

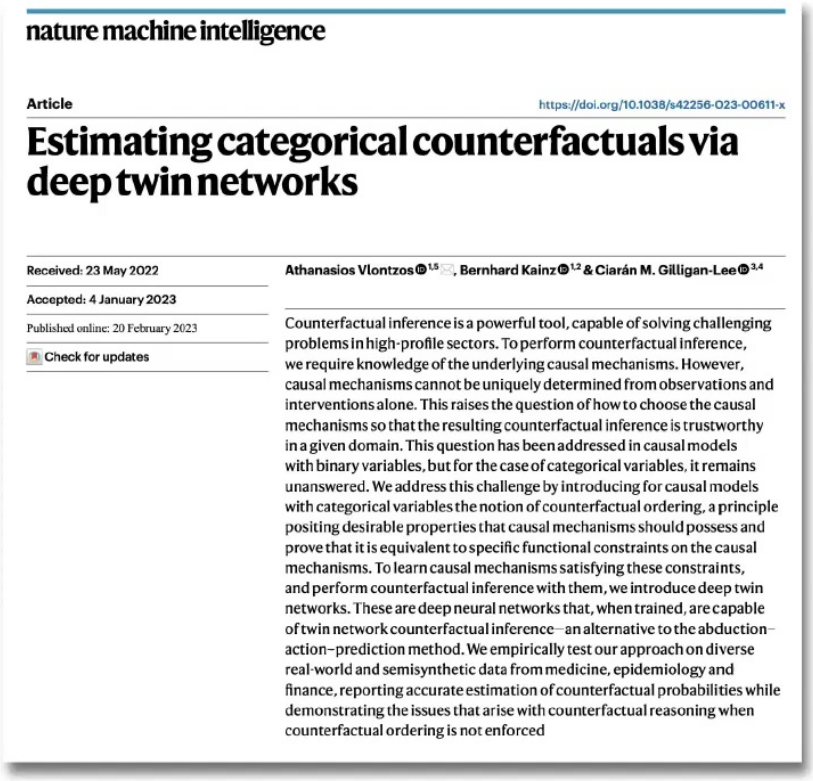
TECHNOLOGY 11 August 2020

By Chris Stokel-Walker

A new medical AI diagnoses people like a doctor would
Reeldeal Images / Alamy

Check out our paper to dive deeper arXiv:1910.06772

Combining it with deep learning allows for personalised decision making from raw data!



nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00611-x

Estimating categorical counterfactuals via deep twin networks

Received: 23 May 2022

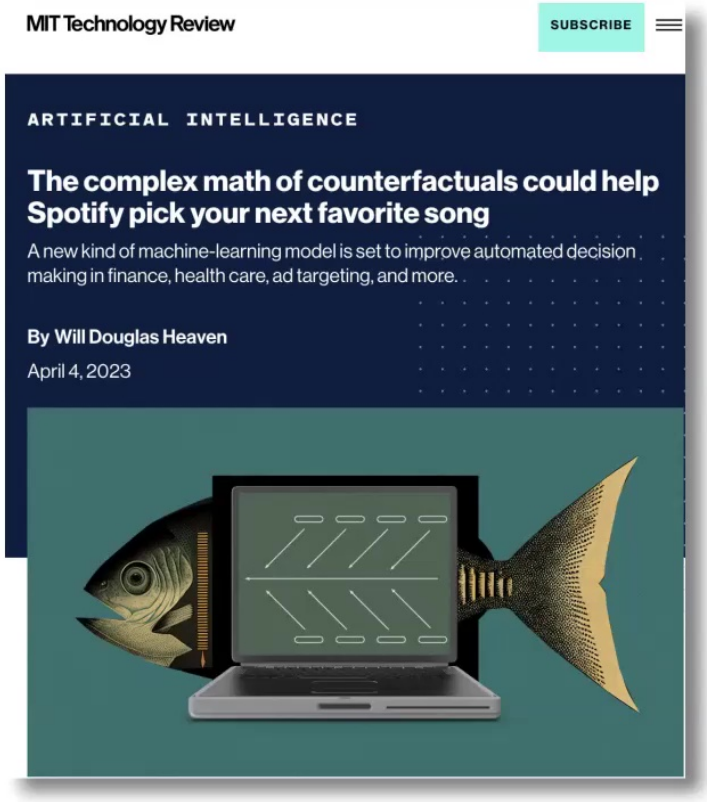
Accepted: 4 January 2023

Published online: 20 February 2023

Check for updates

Athanasios Vlontzos[1,5], Bernhard Kainz[1,2] & Ciarán M. Gilligan-Lee[3,4]

Counterfactual inference is a powerful tool, capable of solving challenging problems in high-profile sectors. To perform counterfactual inference, we require knowledge of the underlying causal mechanisms. However, causal mechanisms cannot be uniquely determined from observations and interventions alone. This raises the question of how to choose the causal mechanisms so that the resulting counterfactual inference is trustworthy in a given domain. This question has been addressed in causal models with binary variables, but for the case of categorical variables, it remains unanswered. We address this challenge by introducing for causal models with categorical variables the notion of counterfactual ordering, a principle positing desirable properties that causal mechanisms should possess and prove that it is equivalent to specific functional constraints on the causal mechanisms. To learn causal mechanisms satisfying these constraints, and perform counterfactual inference with them, we introduce deep twin networks. These are deep neural networks that, when trained, are capable of twin network counterfactual inference—an alternative to the abduction–action–prediction method. We empirically test our approach on diverse real-world and semisynthetic data from medicine, epidemiology and finance, reporting accurate estimation of counterfactual probabilities while demonstrating the issues that arise with counterfactual reasoning when counterfactual ordering is not enforced



MIT Technology Review

SUBSCRIBE

ARTIFICIAL INTELLIGENCE

The complex math of counterfactuals could help Spotify pick your next favorite song

A new kind of machine-learning model is set to improve automated decision making in finance, health care, ad targeting, and more.

By Will Douglas Heaven

April 4, 2023

Check out our paper to dive deeper arXiv:2109.01904

Examples of counterfactuals at Spotify

- **Which Playlists to update:** which playlists Z "need" to be updated?

$$P(Y_{X=\text{update}} = \text{engaged}, Y_{X=\text{no update}} = \text{not engaged} \mid Z)$$

- **New content to enjoy:** If user Z listened to specific content and enjoyed it, which other content would they also have enjoyed?

$$P(Y_{X=\text{new content}} = \text{engage} \mid Y = \text{engage}, X = \text{current content}, Z)$$

And many more….

Conclusion

- Being able to answer causal questions enables actionable decision making
- Lots of new problems to solve if we want to apply causal inference reliably in the real world
- Many more causal inference applications at Spotify beyond what we've discussed today, reach out of you're interested!

ciaranl@spotify

@quantumciaran