

University of California San Francisco



#### Causal Inference, Artificial Intelligence, and Health Research

M. Maria Glymour Professor Department of Epidemiology and Biostatistics University of California, San Francisco

#### Health Research vs Clinical Care

- Lots of (mostly potential) applications in clinical care delivery system
- Medical care is generally not the primary driver of health or of health inequalities
- Primary drivers are factors like:
  - Social disadvantage and opportunities
  - Behaviors
  - Genetics
  - Environmental exposures

#### Outline

- Do health researchers need causal frameworks and AI?
- Have health researchers adopted causal frameworks?
  - DAGs: Two examples where DAGs were key
  - Discovery: no
  - Alternative to trials: no
- Debate and barriers

## What are the most important problems in health research?

- Description
- Prediction
- Prevention
- Treatment

### What are the most important problems in health research?

- Description: Who is sick?
- Prediction: Who is going to get sick?
- Prevention: What can we do to prevent them from getting sick?
- Treatment: What can we do to help them get better?

# Do we need a causal framework to link Al tools to the most important problems in health research?

- Description→ do not need AI or causal framework, though AI could help us get more/better data
   – Caveat: description is hardly ever the real goal.
- Prediction → Just AI could be great. Our prediction models are typically crummy.
  - Caveat: Prediction is usually only valuable if you have causal information in hand. No point being Cassandra.
- Prevention  $\rightarrow$  Nearly all questions are causal.
- Treatment  $\rightarrow$  Nearly all questions are causal.

### Have health researchers adopted a causal framework?

- Health researchers have *always* had a causal framework.
  - Though people get tangled, any physician can distinguish between a prediction model (who should I treat?) and a causal model (how should I treat them?).
  - Epidemiologists usually can too.
- And health research has a long history of careful causal reasoning.
  - Semmelweiss; Goldberger and Sydenstricker (Pellagra)
  - Pose alternative theories of disease etiology
  - Identify testable implications of alternative theories
  - Test and rule out

### Have health researchers adopted DAGs?

- Often, yes, DAGs are simply so convenient, so clarifying, and so accessible that adoption has been widespread.
- Unify disparate concepts.
- Facilitate communication.
- In core textbooks.

Original Article

#### A Structural Approach to Selection Bias

Miguel A. Hernán,<sup>\*</sup> Sonia Hernández-Díaz,<sup>†</sup> and James M. Robins<sup>\*</sup> (Epidemiology 2004;15: 615–625)



#### Has using DAGs been fruitful? Two Applications

#### Correcting a common analytic mistake



American Journal of Epidemiology Copyright © 2005 by the Johns Hopkins Bloomberg School of Public Health All rights reserved Vol. 162, No. 3 *Printed in U.S.A.* DOI: 10.1093/aje/kwi187

When Is Baseline Adjustment Useful in Analyses of Change? An Example with Education and Cognitive Change

M. Maria Glymour<sup>1</sup>, Jennifer Weuve<sup>2</sup>, Lisa F. Berkman<sup>1</sup>, Ichiro Kawachi<sup>1</sup>, and James M. Robins<sup>3</sup>

#### Evaluating the role of survival bias

ORIGINAL ARTICLE

Can Survival Bias Explain the Age Attenuation of Racial Inequalities in Stroke Incidence?

A Simulation Study

Elizabeth Rose Mayeda,<sup>a,b</sup> Hailey R. Banack,<sup>c</sup> Kirsten Bibbins-Domingo,<sup>b,d</sup> Adina Zeki Al Hazzouri,<sup>e</sup> Jessica R. Marden,<sup>f</sup> Rachel A. Whitmer,<sup>g,b</sup> and M. Maria Glymour<sup>b</sup>

#### The Scientific Question

#### does X affect how Y changes over time?



Physical activity -Smoking -Education -

Physical activity  $\rightarrow \Delta$  depressive symptoms

- $\rightarrow$   $\Delta$ lung function
  - $\rightarrow \Delta cognitive function$

Distinguish between biological based human capacities (cognitive function) and measures of each capacity (cognitive test score). 10

#### Brain Function vs Cognitive Test Score

- Distinguish between biological based human capacities (cognitive function) and measures of each capacity (cognitive test score).
- These differ due to:
  - Bad measurement instrument
  - Interviewer effects
  - Context specific effects: bad time of day, bad day of the week, transient changes in mood
  - Practice effects
- Cannot directly assess changes in cognitive function – must rely on changes in cognitive scores.

#### The Practical Question:

How to analyze the data when there are <u>2 measures</u> of Y?



Assume the null hypothesis:

- X has no affect on change in Y.
- Interventions to change X would not benefit changes in Y.
- Would like to know this so we do not bother to spend resources on changing X, but focus on other things to benefit Y.
- Would like an analysis that provides an effect estimate of 0 (or centered around 0) under the null.
- What analysis?

#### The Practical Question:

How to analyze the data when there are <u>2 measures</u> of Y?



#### The Observational Study With a DAG



#### The Observational Study an unbiased analysis

Model of change with no baseline adjustment:  $E[\Delta Y] = \alpha + \beta_1 X$ 

## The Observational Study (typically) **biased** analyses

#### Baseline adjusted model of change: $E[\Delta Y] = \alpha + \beta_1 X + \beta_2 Y_0$

Baseline adjusted model of follow-up Y:  $E[Y_1] = \gamma + \beta_1 X + \gamma_2 Y_0$ 

The Observational Study an unbiased analysis

#### $\mathbf{E}[\Delta \mathbf{Y}] = \alpha + \beta_1 \mathbf{X}$

Disadvantages:

- Statistical power: measurement error in  $\Delta Y$  is large.
- Ceilings: People who start very high or very low cannot change as much as people who start in the middle
  Non-interval scaling: a 1 point decline from a score of 29 is not the same functionally as a 1 point decline from a score of 24.
- •Baseline functioning really does affect change.

The Observational Study (case 1) an unbiased analysis

OLS model of change:  $E[\Delta C] = \alpha + \beta_1 X$ 

Advantages: • Avoids regression to the mean bias, which is often <u>larger</u> <u>than any plausible causal effect</u>

#### What happens in the literature?

- In the education and cognitive change literature, analyses with 2 time points are typically baseline adjusted
- When 3 or more time points are available, mixed effects /growth curve models typically used without baseline adjustment
- The non-baseline adjusted model would center on the same point estimate as a growth curve model under balanced data.
- Many cases the DAG I drew is not right (e.g. RCT) and this implies a different analysis. Key is to draw the DAG to choose the analysis.
- Comes up in many settings, e.g., pharmacogenomics

#### **Second Application**

ORIGINAL ARTICLE

#### Can Survival Bias Explain the Age Attenuation of Racial Inequalities in Stroke Incidence?

A Simulation Study

Elizabeth Rose Mayeda,<sup>a,b</sup> Hailey R. Banack,<sup>c</sup> Kirsten Bibbins-Domingo,<sup>b,d</sup> Adina Zeki Al Hazzouri,<sup>e</sup> Jessica R. Marden,<sup>f</sup> Rachel A. Whitmer,<sup>g,b</sup> and M. Maria Glymour<sup>b</sup>



#### Racial inequalities in stroke

 Qualitative change in racial inequalities in stroke incidence between middle and late life



Benjamin et al., Circulation 2017; Personal communication with Dr. George Howard, REGARDS Study PI, December 2016

#### Alternative explanations are possible...

...because both race groups are automatically covered by the federal Medicare program by age 65, more equal access to preventive services may also contribute to a gap reduction in stroke incidence between the 2 groups. Our findings suggest that the observed disparity in stroke admissions among younger patients may be amenable to expanded medical insurance coverage."

Stroke hospital admission rates per 100 000 population by age and race in South Carolina from 2002 to 2006. Wuwei Feng et al. Stroke. 2009;40:3096-3101

#### ...but have very different implications



# What is driving the qualitative change in racial inequalities in stroke incidence between middle and late life?

- Causal explanation: Improved social conditions for black Americans at older ages
- Selective survival: Among survivors to old age, black Americans represent a more selected, healthier population than white Americans
- Collider-stratification bias *could* occur if unmeasured factors influence mortality and stroke risk



• But is it plausible that the effect could be this big?

#### Simulation study procedures

Express causal structures of interest with DAGs

Specify data-generating process corresponding with each DAG: Pre-specify the "true" age-constant effect of race on stroke incidence

Run 2,000 iterations of sample generation under each causal structure

Estimate the racial disparity in stroke incidence in each age band in each sample

**Quantify magnitude of bias** in each causal structure: Compare the average estimated racial disparity in stroke in each age band with the "true" effect of race on stroke risk

### Hypothetical cohort study of black-white disparities

- Birth cohort of n=20,000 blacks and n=20,000 whites beginning at age 45
- Survival distributions based on US life tables for 1919-1921 birth cohort
- Stroke incidence rates for whites based on REGARDS
- Age-constant effect of black race on stroke incidence
   –stroke rate<sub>black</sub> = stroke rate<sub>white</sub> + 20/10,000 person-years
- U~N(0,1): time-invariant determinant of survival and stroke

Arias, National vital statistics reports 2006; Howard, personal communication Dec 2016

#### Simulation scenario 1: No bias

 No bias scenario: U directly influences stroke risk; no direct effect on mortality risk



#### Simulation scenario 2: Collider Bias

U directly influences stroke risk and mortality risk (HR<sub>stroke</sub>=1.5; HR<sub>mortality</sub>=1.5)



#### Simulation scenario 2: Collider bias with interaction U directly influences stroke risk and mortality risk for blacks; U has no direct effect on mortality for whites



### Simulated survival curves from birth to age 95

Based on U.S. life tables for the 1919-1921 birth cohort Median survival: 65 years whites, 50 years blacks



### Average observed black-white stroke IRD by age band across 2,000 simulated samples



#### Example 2

- •Simulations confirm that selective survival is a plausible explanation for the qualitative change in racial inequalities in stroke incidence
- •Entire structure motivated by the DAG. Critical to understand a core issue of health disparities.

#### Other work heavily motivated by causal framework

- Alzheimer's disease prevention: divergence between RCT results for managing type 2 diabetes (no effect) and observational results (diabetes is a big predictor)
- How to address this? Four types of observational data:
  - Cohorts from small communities or clinics, with small samples, narrow range of diversity but good Alzheimer's disease assessments
  - Cohorts from representative samples, with small samples, good diversity but weak measures
  - Biobanks with completely obscure selection processes, but large samples, mix of good and bad measures
  - Census data covering a whole country, terrible measures
- Data sources cannot be directly pooled

Have health researchers adopted the implications of d-separation and do calculus for causal discovery?

• Absolutely not. Heresy.

 Tiny sliver of exception with instrumental variables approaches, but not generalizing much beyond a narrow set of problems. Have health researchers adopted do calculus as an alternative to RCTs?

• RCTs are still considered gold standard for causal discovery.

#### What are the barriers?

- There's a *lot* of money involved, so algorithms that are not subject to judgment or vulnerable to priors of the researchers are preferable. ITT analysis of RCTs leave less maneuvering room for bias.
- Before advocating to abandon RCTs as the gold standard for evaluating new cancer medications, consider whether you want someone who stands to make hundreds of millions of dollars, or someone whose job depends on demonstrating the drug works, to choose the DAG to guide your observational analysis.

#### What are the barriers?

- Our priors are usually very uncertain
  - Our usual approach relies on even stronger priors but doesn't require us to admit that, thus, it's preferable.
- Our data are truly deeply messy.
  - Outcomes are binary, count, normally distributed, survival etc.
  - Scattershot missingness heavily influenced by unobserved variables
  - Measurement error is nearly universal, and differential measurement error is common. We almost never have a good measure of either exposure or outcome.
  - Sample size typically a problem and in direct tension with measurement quality
- Everyone is too busy writing grants to learn new methods.
- Much of what is 'new' in Pearl et al is completely consistent with long-standing methods or intuition.
- Our most difficult problems are not *solved* by DAGs, they are simply conveniently expressed with DAGs

#### Vandenbroucke et al sparked a small battle in epi with a critique of contemporary causal inference framework

" this theory restricts the questions that epidemiologists may ask and the study designs that they may consider. It also restricts the evidence thatmay be considered acceptable to assess causality, and thereby the evidence that may be considered acceptable for scientific and public health decision making."



International Journal of Epidemiology, 2016, 1–11 doi: 10.1093/ije/dyv341 Original article

Original article

### Causality and causal inference in epidemiology: the need for a pluralistic approach

Jan P Vandenbroucke,<sup>1,\*,†</sup> Alex Broadbent<sup>2,†</sup> and Neil Pearce<sup>3</sup>

#### Davey Smith and Krieger

"Causal inference', in 21st century epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning and using particular representations, such as directed acyclic graphs (DAGs)"



International Journal of Epidemiology, 2016, 1–22 doi: 10.1093/ije/dyw114 Original article

Original article

### The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology

Nancy Krieger<sup>1,\*</sup> and George Davey Smith<sup>2</sup>

<sup>1</sup>Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA and, <sup>2</sup>MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

#### Davey Smith and Krieger likened use of DAGs and counterfactuals to disregard for

"Causal inference', in 21st century epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning and using particular representations, such as directed acyclic graphs (DAGs)"

"One alarming feature of late 20th and current 21st century epidemiological literature on 'causal inference' is the reappearance of previously rebutted causal claims that 'race' is an individual 'attribute' and that it cannot be a 'cause' because is not modifiable'...the problem—one with enormously harmful public health and policy implications that this approach to causal inference and counterfactuals starts at the wrong level, and uses DAGs to bark up the wrong tree and indeed miss the forest entirely."

## Arguments that DAGs restrict scientific domains

Work appears to equate use of DAGs with:

- Turning away from thoughtful evaluation of evidence
  - -Of course, this is not the case. Nearly the opposite.
- Research on structural drivers of inequality

   Also not the case, but divergent perspectives
   on the modifiability criterion muddy the waters

### Why should we try to overcome those barriers?

- –Unnecessary RCTs are unethical and expensive.
- -We are leaving information on the table
- New data sources make some tools that were previously untenable more viable, e.g., instrumental variables
- -Slow uptake is delaying scientific progress on prevention and treatment of disease

#### How Can Al Accelerate Health Research

- Our most important questions are causal and need fairly precisely defined independent variables to justify and guide public health actions.
- To estimate those causal models, we need:
  - Prediction models for confounding control
  - Prediction models for effect heterogeneity (Athey, Wager, et al)
  - Hypothesis generation?
  - Data scraping and modeling to improve assessments of outcomes without clinical diagnoses (ICD codes are a terrible proxy for health)
- Excellent trainees and collaborators



Thanks to: ER Mayeda J Weuve JM Robins LF Berkman I Kawachi MC Power K Bibbins-Domingo PhD Program in Epidemiology and Translational Science



NIA: R01AG057869