The Alan Turing Institute

Intro to Causal Inference + Counterfactual Prediction

www.turing.ac.uk
@turinginst
@theturinginst

#TMCFTURING

PETER WG TENNANT PHD
ROMIN PAJOUHESHNIA PHD
Intro to Causal Inference

Peter WG Tennant
University of Leeds, UK

#TMCFTURING

The Alan Turing Institute
LEEDS Institute for Data Analytics
UNIVERSITY OF LEEDS
“I sit in a dark room and project code straight to my face while solving complicated problems. This helps me to immerse myself in it and "feel" the code.”
WHAT IS DATA SCIENCE?

DATA SCIENCE

STATISTICS

COMPUTER SCIENCE

DOMAIN EXPERTISE
Science is about building knowledge and understanding

Data science is about ‘gaining insights’ and ‘extracting meaning’ from data

Most data science activity can be divided into three scientific tasks, each with different methods and philosophies

Description (& visualisation)

- Focussed on summarising, describing, &/or visualising features of interest
- Data driven - involves simple calculations & unsupervised learning

Questions

- What happened?
- Who was affected?
- What was occurrence of Y in people with X?

Example

- Occurrence and spread of COVID-19
Prediction (AKA classification and regression)

- Focussed on pattern recognition and forecasting
- Data driven – involves statistical modelling and supervised learning

Questions

- What will happen?
- Who will be affected?
- Are people with X are more likely to have Y?

Examples

- Screening for COVID-19 with symptoms or CT scan
- Predicting prognosis or severity of infection
Causal inference (AKA counterfactual prediction)

- Focussed on understanding
- **NOT data driven** – involves fusion of external knowledge with statistical modelling and supervised learning

Questions

- What will happen if…?
- Why were they affected?
- If we changed \( X \), how would it change \( Y \)?

Examples

- Effect of opening/closing schools on infection spread
- Risks / benefits of ventilation
We learn about cause-and-effect from the moment we’re born

From observing how things interact, and what happens when we do things
This extends beyond what we can see & do, to whatever we can imagine!

Our ability to ask ‘what might have been?’ and ‘how things could be different?’ has led to us changing our world beyond our ancestors’ imaginations.
EASILY FOOLED

https://www.deviantart.com/butisit/art/Checker-shadow-illusion-263331875

#TMCFTURING

https://www.newslinq.com/24-photos-need-really-look-understand-v1/
This is especially true once we’re faced with questions and events that don’t fit our childhood model of cause-and-effect…

A MIND FOR DETERMINISM

HEALTHY + VIBRIO CHOLERAE → UNHEALTHY

✓ Deterministic
This is especially true once we’re faced with questions and events that don’t fit our childhood model of cause-and-effect…

A MIND FOR DETERMINISM

Probabilistic

HEALTHY

SMOKING

HEALTHY

× Probabilistic
Statistics & probability have resolved this problem

What’s **unknowable** for an individual… can be **predictable** for a group!
With **randomisation**, this provides a potent way to estimate causal effects that we’d have little hope with experience and intuition alone.
In health and medical research, **randomised controlled trials** embraced with impassioned fervour, and canonised beyond all other forms of evidence.

**RANDOMISATION**

RCTs indexed in PubMed by year:

- **Level 1**: Systematic review of RCTs
- **Level 2**: Systematic review of nonrandomized trials
- **Level 4**: Single RCT
- **Level 5**: Single nonrandomized trial
- **Level 6**: Systematic review of correlational / observational studies
- **Level 7**: Single correlational / observational study
- **Level 8**: Systematic review of descriptive / qualitative / psychologic studies
- **Level 9**: Single descriptive / qualitative / psychologic studies
- **Level 10**: Opinions of authorities, expert committees

The Alan Turing Institute
Non-experimental studies have a poor record for causal inference

12 RCTs studied 52 claims from non-experimental data

- 0 replicated
- 5 found opposite!

<table>
<thead>
<tr>
<th>ID no.</th>
<th>Pos.</th>
<th>Neg.</th>
<th>No. of claims</th>
<th>Treatment(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>Vit E, beta-carotene</td>
<td>NEJM 1994; 330: 1029–1035</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>Hormone Replacement Ther.</td>
<td>JAMA 2003; 289: 2651–2662, 2663–2672, 2673–2684</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Vit E, beta-carotene</td>
<td>JNCI 2005; 97: 481–488</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>Vit E</td>
<td>JAMA 2005; 293: 1338–1347</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>Low Fat</td>
<td>JAMA. 2006; 295: 655–666</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>Vit D, Calcium</td>
<td>NEJM 2006; 354: 669–683</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Folic acid, Vit B6, B12</td>
<td>NEJM 2006; 354: 2764–2772</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Low Fat</td>
<td>JAMA 2007; 298: 289–298</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>Vit C, Vit E, beta-carotene</td>
<td>Arch Intern Med 2007; 167: 1610–1618</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>Vit C, Vit E</td>
<td>JAMA 2008; 300: 2123–2133</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>Vit E, Selenium</td>
<td>JAMA 2009; 301: 39–51</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>HRT + Vitamins</td>
<td>JAMA 2002; 288: 2431–2440</td>
</tr>
</tbody>
</table>

Totals 0 5 52

Young & Karr 2011 Significance, 116:120, 2011
**Teach:** Correlation ≠ causation

**Per capita cheese consumption**

correlates with

**Number of people who died by becoming tangled in their bedsheets**

- **Discourage:** Causal inference from non-experimental data

https://www.tylervigen.com/spurious-correlations
But: We are programmed to infer causality…

And: Spurious correlations have limited scientific and practical interest

So: People tend to infer causality regardless
5. Discussion

Among 14-year olds living in the UK, we found an association between social media use and depressive symptoms and that this was stronger for girls than for boys. The magnitude of these associations reduced when potential explanatory factors were taken into account, suggesting that experiences of online harassment, poorer sleep quantity and quality, self-esteem and body image largely explain observed associations. There was no evidence of differences for girls and boys in hypothesised pathways between social media use and depressive symptoms. **Findings are based largely on cross sectional data and thus causality cannot be inferred.**

Our findings add weight to the growing evidence base on the potential pitfalls associated with lengthy time spent engaging on social media. These findings are highly relevant to current policy development on guidelines for the safe use of social media and calls on industry to more tightly regulate hours of social media use for young people [[10], [11]].

Clinical, educational and family settings are all potential points of contact whereby young people could be encouraged to reflect not only on their social media use but also other aspects of their lives including online experiences and their sleep patterns. For instance, in the home setting all family members could reflect on patterns of use and have in place limits for time online, curfews for use and the overnight removal of mobile devices.

Kelly Y et al. Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. EClinicalMedicine 2019.
Increased mortality associated with weekend hospital admission: a case for expanded seven day services?

Nick Freemantle and colleagues discuss the findings of their updated analysis of weekend admissions and the implications for service design.

Freemantle et al 2015
BMJ 5;351
Quiet hospitals kill, but mindless union bashing can give us a nasty injury

Jeremy Hunt has every right to question NHS consultants’ working practices but there are questions they should be allowed to ask him too.

Two deaths possibly linked to 'Hunt effect', study suggests

Research suggests some patients may be avoiding going to hospital at the weekend because of health secretary’s statements about inadequate NHS staffing levels.
STOP NON-EXPERIMENTAL RESEARCH?

Physical
(E.g. Repetitive strain)
Chemical
(E.g. Asbestos)
Biological
(E.g. Influenza)
Cultural
(E.g. Childhood diet)
Economic
(E.g. Poverty)
Environmental
(E.g. Pollution)
Political
(E.g. Austerity)
Behavioural
(E.g. Smoking)
Personal
(E.g. Self-esteem)

Psychological
Social
The Alan Turing Institute
Accept & admit our causal ambitions

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Data Are Not Enough—Hurray For Causality!

The C-Word: The More We Discuss It, the Less Dirty It Sounds

Start With the “C-Word,” Follow the Roadmap for Causal Inference

CHANGE NON-EXPERIMENTAL RESEARCH!
Upgrade our **epistemological (philosophical) machinery**

"How can we infer causal relations from observations?"

\[ p(X = x) = \prod_i p(X_i = x_i | PA_i = pa_i) \]

**Philosophy**  
**Maths**  
**Data science**
Machine learning is excellent at finding patterns in complex data > **prediction**

Causal inference requires identifying and estimating **counterfactuals** (i.e. events that did **NOT** happen).

In non-experimental data, machine learning can’t do **causal inference** because you **CAN’T** learn what’s not there!

Need external knowledge of ‘**data generating process**’; the story behind how the data came into being.
JOHN SNOW, LEGEND, DATA SCIENTIST

‘BROAD STREET’ PUMP

Based on a true story.

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 Aug</td>
<td>Cholera outbreak in Soho.</td>
</tr>
<tr>
<td>5 Sep</td>
<td>Snow maps where deaths are occurring.</td>
</tr>
<tr>
<td>6 Sep</td>
<td>83% of deaths drank from Broad Street pump.</td>
</tr>
<tr>
<td>7 Sep</td>
<td>Snow meets parish guardians to argue for pump closure.</td>
</tr>
<tr>
<td>8 Sep</td>
<td>Snow removes handle &gt; Cholera outbreak ends.</td>
</tr>
</tbody>
</table>
Snow observed the **outcome** of what happened when the pump was ‘closed’ **(factual)**
Didn’t observe the potential outcome that would have happened if the pump had been – counter to fact - left ‘open’ (counterfactual)

Pump left open
- We can never know the potential outcome for a counterfactual exposure!
- For each ‘unit of analysis’ we can only observe one potential outcome
- This is known as the fundamental problem of causal inference
Instead we must *estimate* the potential outcome for the counterfactual exposure from exchangeable units of analysis.
ESTIMATING POTENTIAL OUTCOMES

GIVE MAGIC BEANS
Problem:
- Units (e.g. people) are very different
- Even the same units can respond differently at different times

Not exchangeable units of analysis!
POOR COUNTERFACTUALS

GIVE MAGIC BEANS
Poor estimate of potential outcome!
- We therefore have to **identify** \( (sub)groups \) of units that were **exchangeable** at the time of exposure.

- We can then estimate the **average causal effect** by **comparing** the **outcomes** between these \( (sub)groups \).
AVERAGE CAUSAL EFFECTS

GIVE MAGIC BEANS
AVERAGE CAUSAL EFFECTS
The easiest way to do this is through randomisation.
Without randomisation, you have to identify and compare like-for-like subgroups.
Without randomisation, you have to identify and compare like-for-like subgroups.
Without randomisation, you have to identify and compare like-for-like subgroups.

Need to identify, measure, & control for *every* variable that (potentially) *causes* both the exposure and outcome (**confounder**).
Causal inference methods – like Judea Pearl’s ‘Structural Causal Model’ – provide a mathematical and philosophical framework for considering causal effects drawing on:

- Probability theory
- Counterfactual reasoning
- Graphical model theory

Judea Pearl, Computer scientist, philosopher, Turning Prize winner

2000

2009

2016

2018
**Key feature:** Formally identifying your effect before analysis!

**ESTIMAND**
What you seek
E.g. The true difference in Y due to exposure

**ESTIMATOR**
How you will get there
E.g. Your regression model

**ESTIMATE**
What you get
E.g. the estimated difference in Y from model coefficient
Causal diagrams – such as directed acyclic graphs – encode counterfactual & probabilistic reasoning into diagrammatic form

First benefit: helps to identify which variables are confounders and need controlling to provide exchangeable comparisons

We must do this with our outside knowledge, the machine/software cannot!
Directed Acyclic Graphs (DAGs) are nonparametric representations of the (hypothesised) causal relationships between variables.

- Relationships between variables (‘nodes’) are represented by arrows (‘arcs’) creating paths between them.
- Paths can be causal or confounding; open or closed.
- Open paths transmit correlations, closed paths do not.
To estimate the causal effect of **C** on **F** (our ‘focal relationship’):

- We want all *causal paths* open
- And all *confounded paths* closed
To estimate the causal effect of \( C \) on \( F \) (the ‘focal relationship’):
- We want all causal paths open
- And all confounded paths closed
- This means controlling for all confounders but no mediators
Example: Total causal effect of $X_3$ on $Y$:

- Model should include confounders ($X_1$, $X_2$) [and competing exposures ($X_5$)]
- $Y \sim X_3 + X_1 + X_2 [+ X_5]$
This model was constructed to estimate the effect of $X_3$ on $Y$

We should not interpret the other covariates ($X_1, X_2, X_5$), because they will not be appropriate controlled.

E.g. The coefficient on $X_1$ is **NOT** the total causal effect of $X_1$ on $Y$, because the inclusion of $X_3$ means the causal path $X_1 \rightarrow X_3 \rightarrow Y$ will be closed.

$$Y \sim X_3 + X_1 + X_2 + X_5$$
The tradition of including all ‘predictors’ of our outcome (Y) in a single model, and interpreting the coefficients (X_1, X_2, X_3, X_4, X_5) as ‘mutually adjusted’ effects has been dubbed the ‘Table 2 Fallacy’.

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich* and Sander Greenland

* Correspondence to Dr. Daniel Westreich, Department of Obstetrics and Gynecology, Duke Global Health Institute, Duke University, DUMC 3967, Durham, NC 27710 (e-mail: daniel.westreich@duke.edu).

Initially submitted January 13, 2012; accepted for publication October 11, 2012.

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

• DAGs have also been revolutionary in helping us understand **conditional dependencies** – non-causal associations introduced by conditioning.

• Conditioning on \( C \) would transmit a **non-causal association** between \( X \) and \( Y \).
**Example:** The Sunglasses Revelation

- British people typically wear sunglasses when the **sun is shining**.
- They sometimes also wear them when they are **very hungover**.

On average, suppose hangovers occur equally all year round, regardless of weather.
**Example:** The Sunglasses Revelation

- You see someone **wearing sunglasses** on a **cloudy day**

- What does this tell you about their likelihood of being **hungover**?
Practice of Epidemiology

The Birth Weight “Paradox” Uncovered?

Sonia Hernández-Díaz¹,², Enrique F. Schisterman³, and Miguel A. Hernán¹

¹ Department of Epidemiology, Harvard School of Public Health, Boston, MA.
² Sloane Epidemiology Center, Boston University, Boston, MA.
³ Epidemiology Branch, National Institute of Child Health and Human Development, Bethesda, MD.

Received for publication February 7, 2005; accepted for publication January 23, 2006.

THE BIRTHWEIGHT PARADOX

Smoking → Low birth weight → Infant death
In our model, we recognised the coefficient on $X_1$ was **NOT** the total causal effect of $X_1$ on $Y$, but we might think it gave a meaningful (direct) causal effect.

In fact, the apparent effect will be **collider-biased** by all other mutual causes of $X_3$ and $Y$ (**mediator-outcome confounders**)

This can **reverse** the apparent direction of effect (**reversal paradox**)!
Causal inference is one of the three core tasks of data science.

Unlike description and prediction, it requires external contextual knowledge.

Traditional data-driven approaches have a poor record at causal inference.

Joining the causal revolution means embracing our causal ambitions, & recognising the need for a completely different approach.

This starts with explicitly identifying what we want to know before estimating.

Causal diagrams are a simple & transparent aid to causal reasoning, & are helpful for understanding many issues in non-experimental data.
Causal inference requires external knowledge → careful thought

Data scientists need more time to THINK before they DO!
Differences between prediction and causal inference


Using DAGs in your research


Causal Inference Book