TABLE 2 FALLACY

Or why interpretation needs more than transparency

Peter WG Tennant PhD

University Academic Fellow (University of Leeds)

Fellow (Alan Turing Institute)

@PWGTennant

The Alan Turing Institute

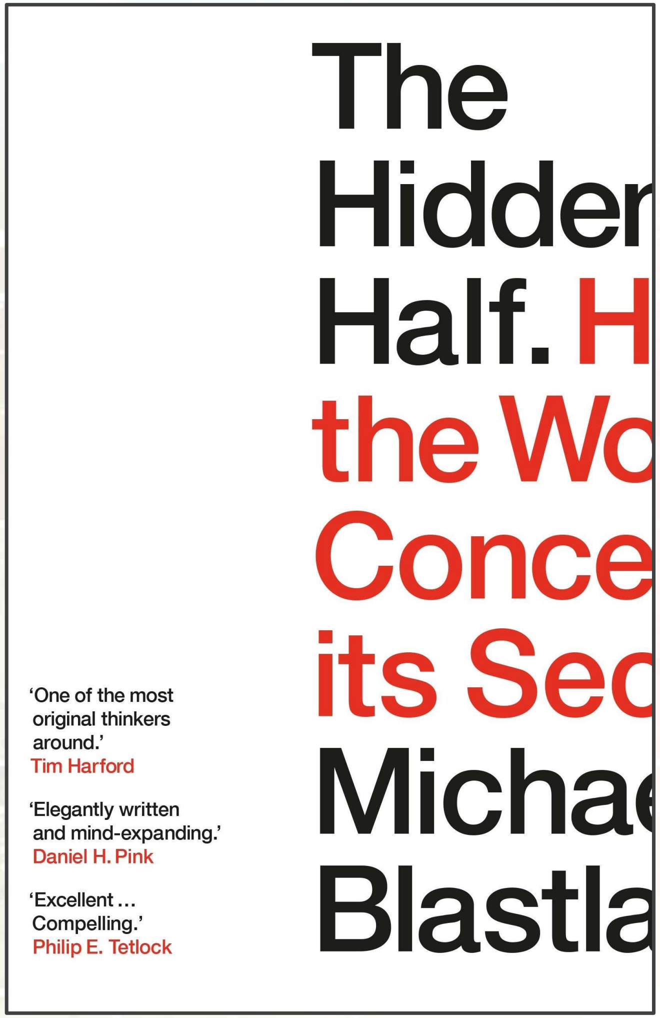
LEEDS Institute for Data Analytics

UNIVERSITY OF LEEDS

THE RISE OF ALGORTHIMS

- Algorithms are increasingly used in our world for **pattern recognition**, **profiling** & **decision making**
  - Speech & language recognition
  - Classifying & labelling images
  - Screening job applicants
  - Approving insurance & loans
  - Diagnosing diseases
  - Informing treatments
  - Advertisements & offers
  - What you see on social media

PROBLEMS WITH ALGORITHMS



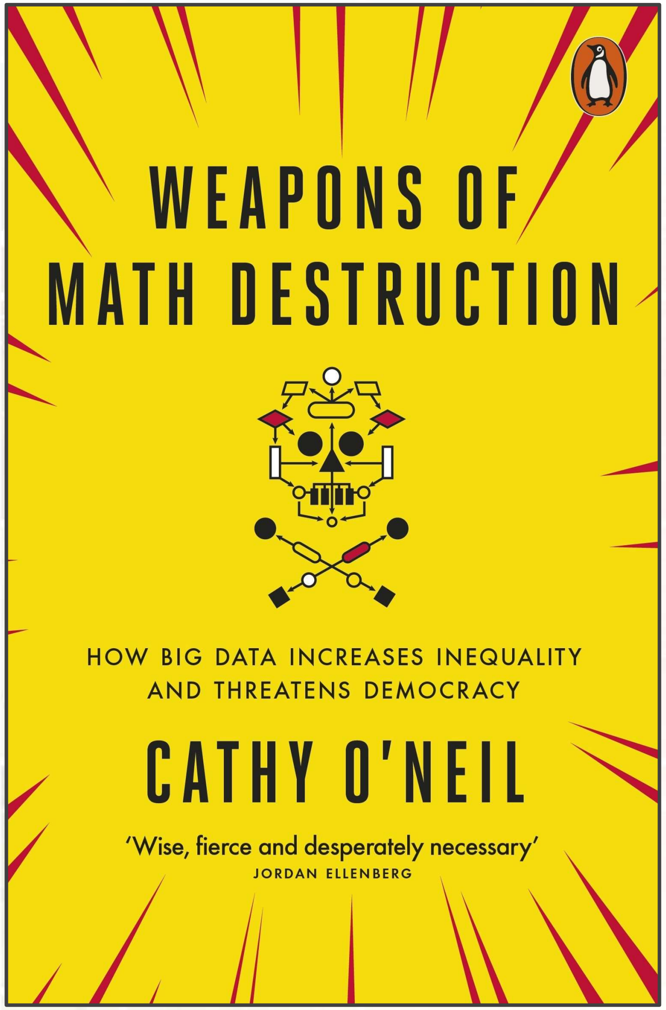
- They are fundamentally unsuited to individual-level predictions/decisions
- **Wilkinson et al 2020 – *Lancet Digital Health***

Time to reality check the promises of machine learning-powered precision medicine

Jack Wilkinson, Kellyn F Arnold, Eleanor J Murray, Maarten van Smeden, Kareem Carr, Rachel Sippy, Marc de Kamps, Andrew Beam, Stefan Konigorski, Christoph Lippert, Mark S Gilthorpe, Peter W G Tennant

"Although statistics—and hence machine learning—is excellent at helping us to understand and compare probabilities between groups, it is fundamentally unable to tell us what will happen to an individual. The power of statistics is precisely that it can describe and predict partly random events over large numbers of people."

PROBLEMS WITH ALGORITHMS



- They are (primarily) **data-driven**
  - They are excellent at identifying and utilising **patterns** and **associations** within data
  - They have no *understanding* what these patterns and associations *mean*
  - They are *excellent* at encoding – and magnifying – **prejudice** and '**bias**' within the data
- Many **data-driven algorithms** are hence morally and socially regressive
- Many are: '*weapons of math destruction*'

THE A-LEVEL RESULTS ALGORITHM

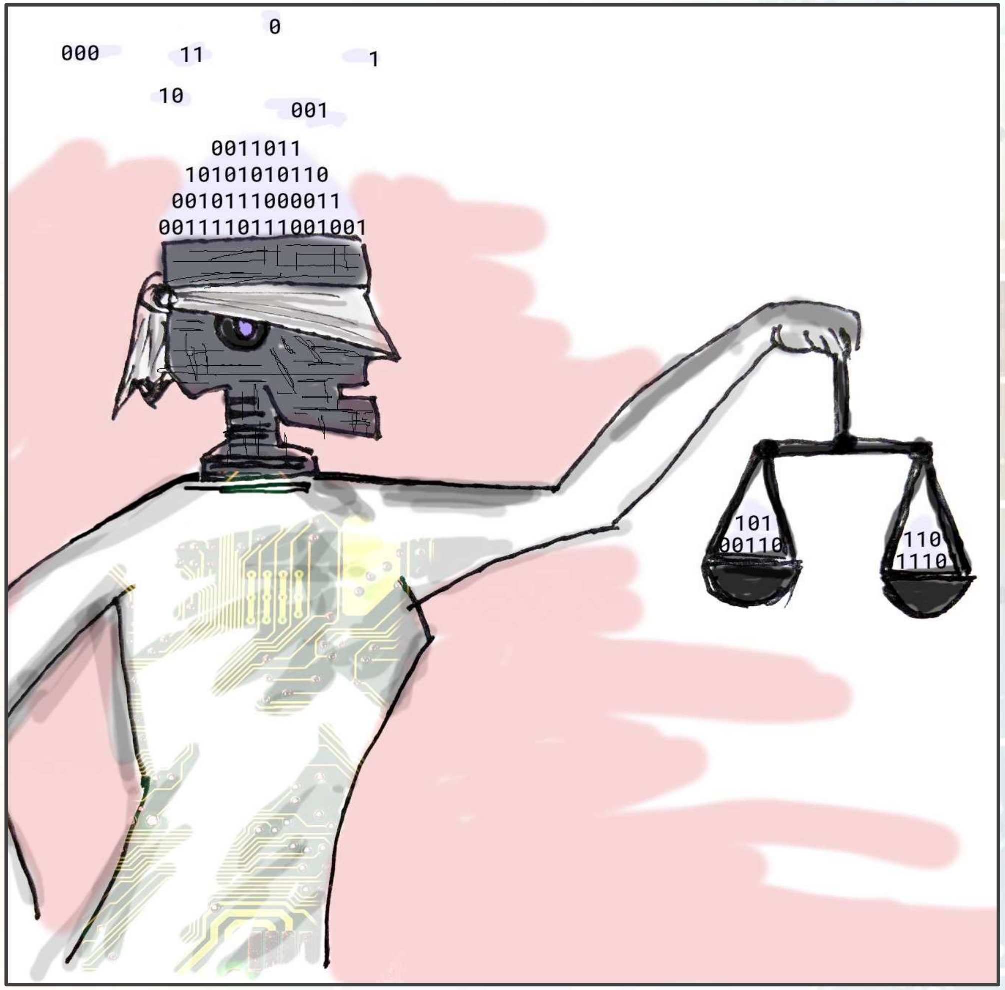
- Problems exemplified by Ofqual’s A-level results algorithm
- Output performance seemed reasonable, but failed at individual level
- Exposed **existing bias** in society:
  - Students from **less advantaged schools** were **systematically downgraded**
- Much was written about this disastrous algorithm
  - But – had exams gone ahead – the same disadvantage would have occurred!

$$P_{kj} = (1-r_j)C_{kj} + r_j(C_{kj} + q_{kj} - p_{kj})$$

ALGORITHMIC JUSTICE

If desire **equity** and **justice**, then **must:**

- **Design** algorithms to:
  - Incorporate **agnostic / fair features**
  - Ignore/correct **prejudiced / unfair features**
- **Scrutinize** algorithms to:
  - Maximise **fairness**
  - Reduce **unintended consequences**



EXPLAINABILITY & INTERPRETABILITY

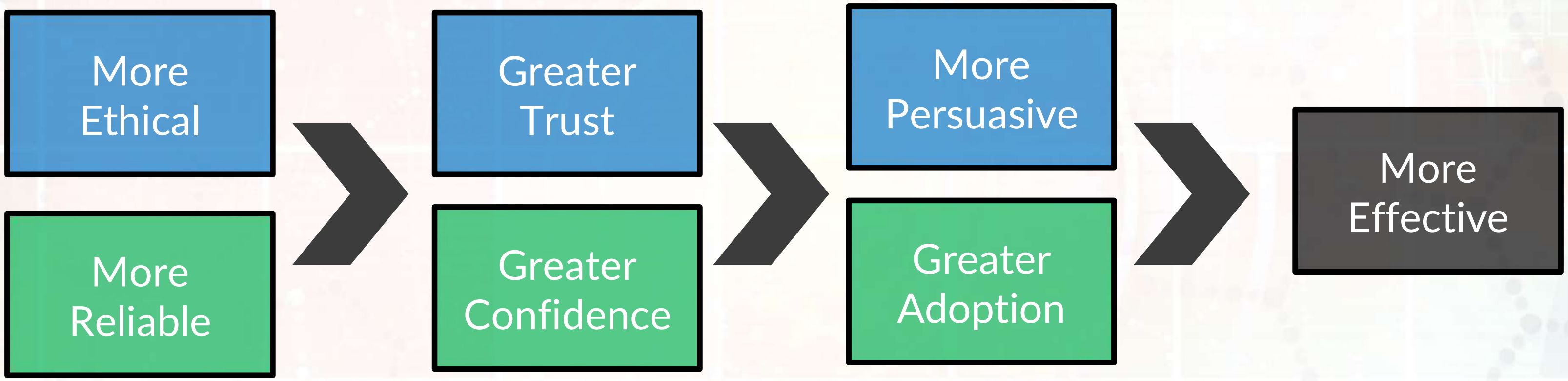
- One way to enhance algorithmic justice would be to design more **explainable** &/or **interpretable** algorithms
- There is some debate over definitions, but:

EXPLAINABILITY	INTERPRETABILITY
*"(The) assignment of **causal** responsibility"* **- Joseph & Joseph, 1996**	*"The degree to which an observer can understand the cause of a decision"* **- Brian & Cotton, 2017**

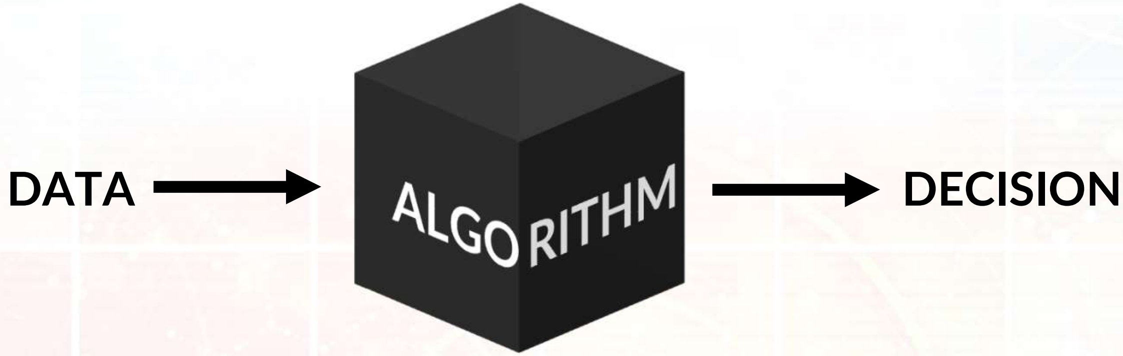
- **Broadly:** An explainable / interpretable algorithm is one where the ***reasons*** for a decision can be queried and explained in a way that makes sense to humans.

MORE EFFECTIVE ALGORITHMS

- **Explainable** and **interpretable** algorithms also promise greater **reliability**
- **Together:**



MORE TRANSPARENT ALGORITHMS



- **Black-box algorithms** are opaque models where it is not clear *why* decisions are being made
- These are *not natively compatible* with **interpretability**
  - They may **encode** many **prejudices**
  - Their '**features**' may not transport well
- There is hence a drive for more ***transparent*** algorithms

MORE TRANSPARENT ALGORITHMS

European Parliament

A governance framework for algorithmic accountability and transparency

Contact | Stay Con

FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

ABOUT THE FTC | NEWS & EVENTS | ENFORCEMENT | POLICY | TIPS & ADVICE

Home » Tips & Advice » Business Center » Business Blog » Using Artificial Intelligence and Algorithms

Using Artificial Intelligence and Algorithms

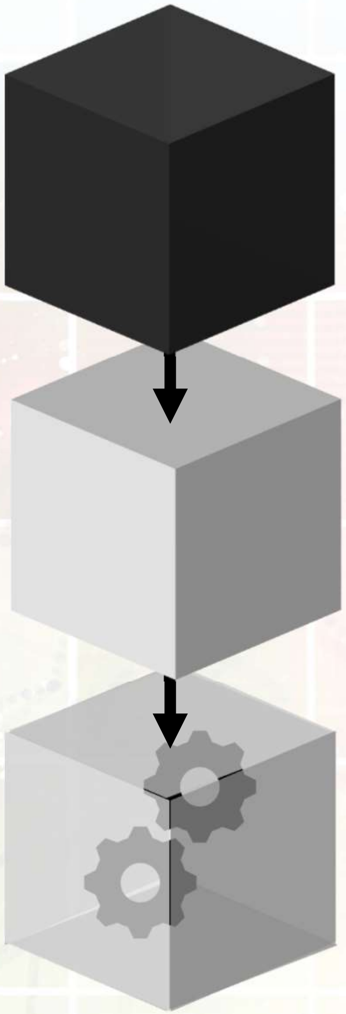
By: Andrew Smith, Director, FTC Bureau of Consumer Protection | Apr 8, 2020 9:58AM

SHARE THIS PAGE

TAGS: Bureau of Consumer Protection | Consumer Protection | Privacy and Security | Consumer Privacy | Credit Reporting | Data Security | Tech

Headlines tout rapid improvements in artificial intelligence technology. The use of AI technology – machines and algorithms – to make predictions, recommendations, or decisions has enormous potential to improve welfare and productivity. But it also presents risks, such as the potential for unfair or discriminatory outcomes or the perpetuation of existing socioeconomic disparities. Health AI offers a prime example of this tension. Research recently published in *Science* revealed that an algorithm used with good intentions – to target medical interventions to the sickest patients – ended up funneling resources to a healthier, white population, to the detriment of sicker, black patients.

MORE TRANSPARENT ALGORITHMS



- **Symbolic metamodeling** is an approach to increasing transparency
  - A '***white-box model***' is produced to **mimic** the performance of a **black-box algorithm**
  - Aside: Surely '***transparent box***' would be better?!
- **Transparency** *seems* like an important step to **interpretability**
  - Can extract **features** and **relative importance** towards decision?

> **If you use algorithms to assign risk scores to consumers, also disclose the key factors that affected the score, rank ordered for importance**. Similar to other algorithmic decision-making, scores are based on myriad factors, some of which may be difficult to explain to consumers. For example, if a credit score is used to deny someone credit, or offer them less favorable terms, the law requires that consumers be given notice, a description of the score (its source, the range of scores under that credit model), and at least four key factors that adversely affected the credit score, listed in the order of their importance based on their effect on the credit score.

Source: https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms

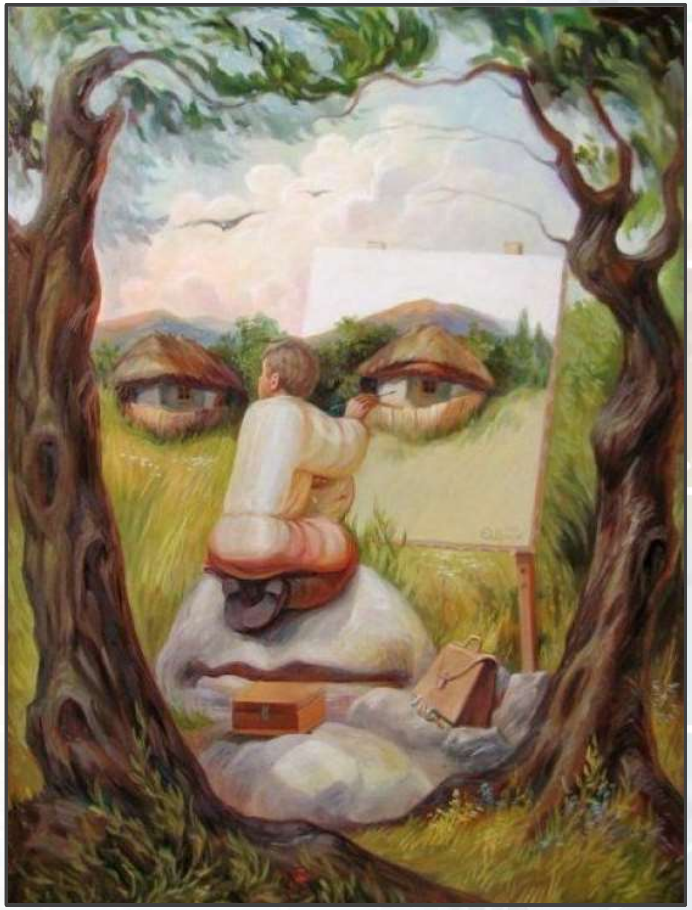
TRANSPARENCY ≠ INTERPRETABILITY

- Being able to **identify** and **describe** features does **not** mean those features are **interpretable**!

Artist: Oleg Shuplyak

TRANSPARENCY ≠ INTERPRETABILITY

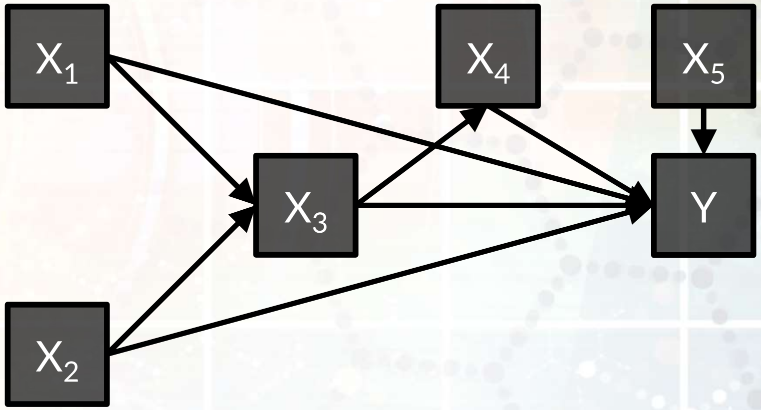
- Being able to **identify** and **describe** features does **not** mean those features are **interpretable**!
- In a **data-driven prediction model**, the features:
  - Have **no real-world meaning**
  - Represent an ***obscure*** combination of variables inside *and outside* the model!
  - Should **not** be ranked for ***relative importance***



Artist: Oleg Shuplyak

INTERPRETABILITY REQUIRES A CAUSAL MODEL

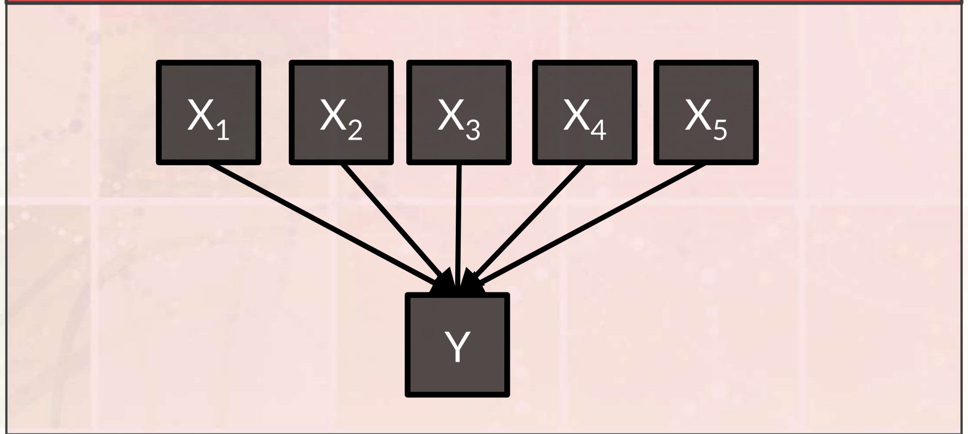
- **Interpretability** & **explainability** requires understanding ***causes*** and ***effects***
- This **cannot** be **determined** from the **data alone**
- Requires understanding of the **data generating mechanism**:
  - Context
  - Data lineage
  - Sampling and selection
  - Meaning / relationship between variables



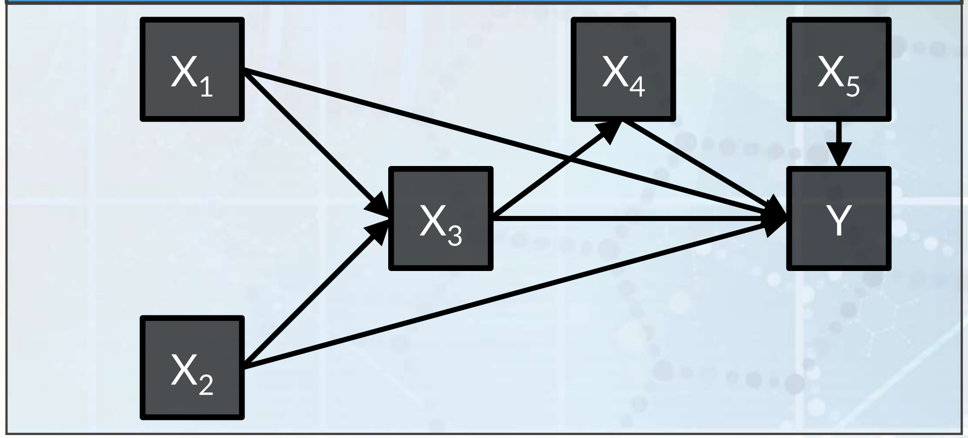
INTERPRETABILITY REQUIRES A CAUSAL MODEL

- Knowledge of the **data-generating mechanism** has to be provided by **external theory** and **understanding**;
  - I.e. a **causal model**!
- No **software/algorithm** can (currently) understand this
  - **Data-driven predictive models** cannot be interpreted – however transparent

HOW THE ALGORITHM SEES IT



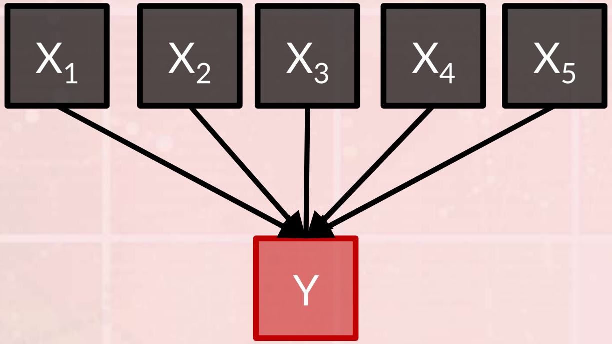
HOW NATURE CREATED IT



PREDICTIVE VS CAUSAL MODELLING

PREDICTIVE MODEL

- **Outcome**-focused

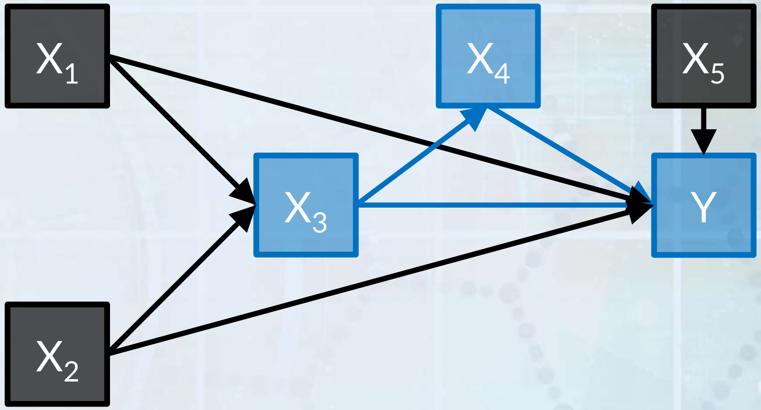


X_i is ____ Y

'correlated with'
'a predictor of'
'associated with'

CAUSAL MODEL

- **Effect**-focused



The ____ of X_2 on Y is…

'total causal effect'
'direct causal effect'

PREDICTIVE VS CAUSAL MODELLING

PREDICTIVE MODEL

- **Aim: Predict values of outcome**
- **Maximise: Variance ‘explained’ (R^2)**
- **Covariate selection focused on:**
  - Balancing **precision** & **parsimony**
  - **Availability** of variables
  - **Maximising: Joint information**
- **Coefficients: Uninterpretable**
- **Automation: Favoured**

CAUSAL MODEL

- **Aim: Estimate a causal effect**
- **Maximise: Accuracy of estimate**
- **Covariate selection focused on:**
  - External **knowledge** & **judgement**
  - **Role** of variables
  - **Minimizing: confounding** & **selection bias**
- **Coefficients: Interpretable**
- **Automation: Not possible**

EXAMPLE 1 - TRANSPARENT BUT UNINTERPRETABLE

Bald men at higher risk of severe case of Covid-19, research finds

Researchers suggested that baldness should be considered a risk factor, dubbing it the 'Gabrin sign'

By Jennifer Rigby
4 June 2020 • 8:33pm

Source: https://www.telegraph.co.uk/global-health/science-and-disease/bald-men-higher-risk-severe-case-covid-19-research-finds/

Received: 7 April 2020 | Accepted: 14 April 2020
DOI: 10.1111/jocd.13443

LETTER TO THE EDITOR

JCD Journal of Cosmetic Dermatology
WILEY

A preliminary observation: Male pattern hair loss among hospitalized COVID-19 patients in Spain – A potential clue to the role of androgens in COVID-19 severity

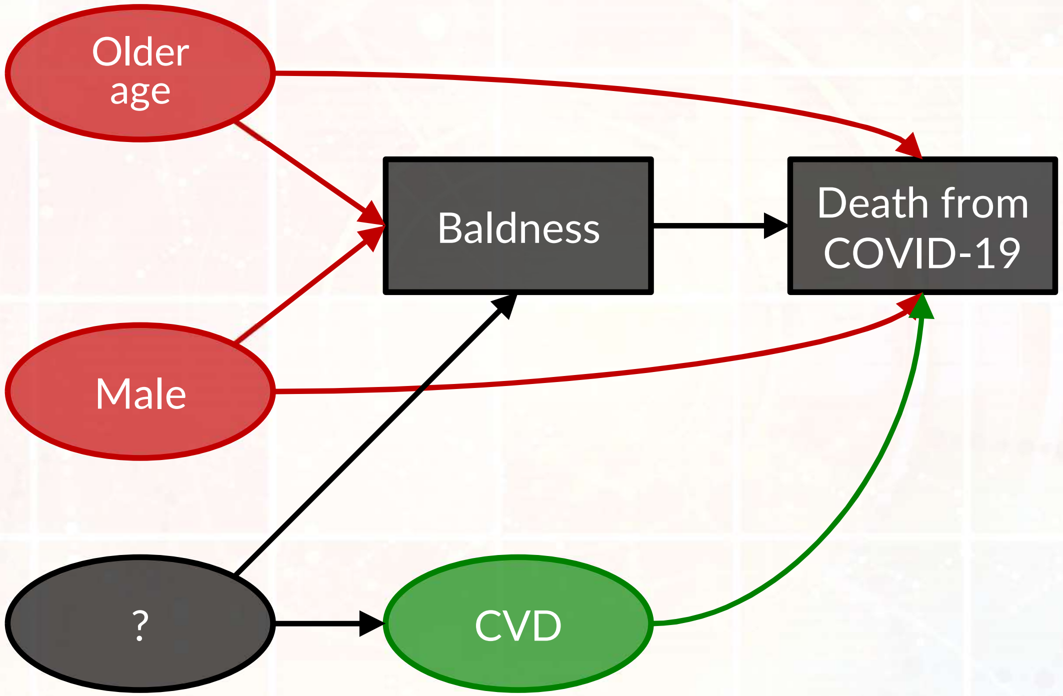
Abstract
A preliminary observation of high frequency of male pattern hair loss among admitted COVID-19 patients and suggest that androgen expression might be a clue to COVID-19 severity.

dependent on genetic variants found in the androgen receptor gene located on the X chromosome. We hypothesized that males with AGA are more likely to be hospitalized for COVID-19 complications compared to controls. To explore this potential association, we conducted a preliminary observational study of the prevalence of AGA patients among hospitalized COVID-19 patients at two Spanish tertiary hospitals between March 23 and April 6, 2020, the diagnosis of

Source: Goren *et al* 2020 - *J Cosmet Dermatol*

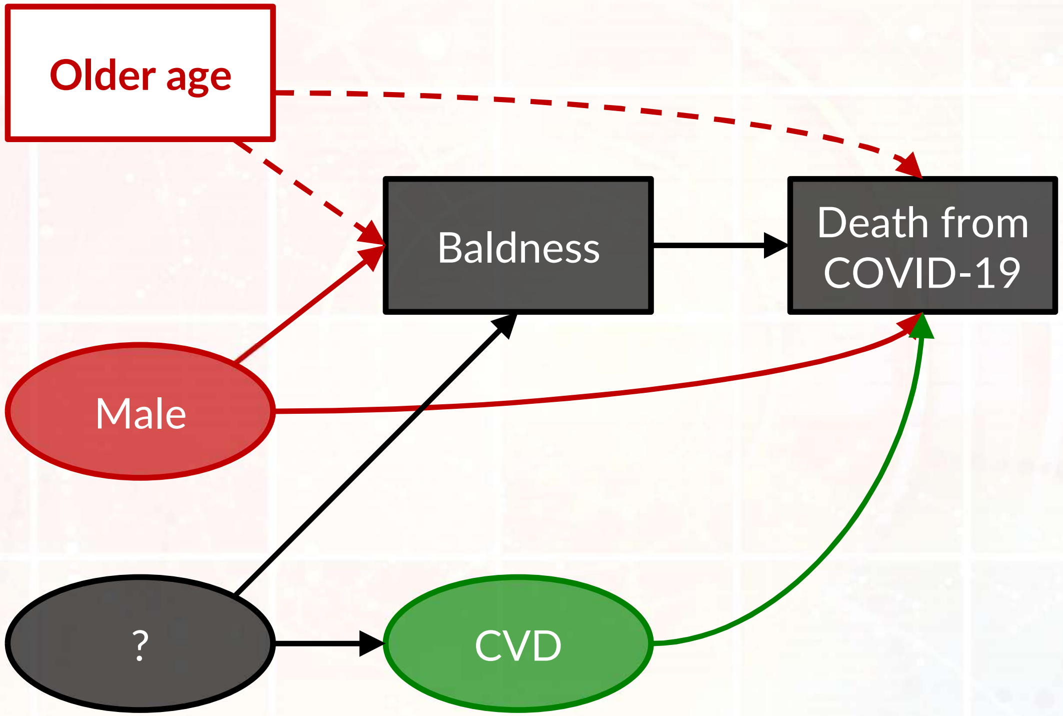
EXAMPLE 1 - TRANSPARENT BUT UNINTERPRETABLE

- How ***important*** is baldness?
- It depends…



OMITTED VARIABLE BIAS?

- How ***important*** is baldness?
- It depends…



OMITTED VARIABLE BIAS?

Article

Factors associated with COVID-19-related death using OpenSAFELY

https://doi.org/10.1038/s41586-020-2521-4
Received: 15 May 2020
Accepted: 1 July 2020
Published online: 8 July 2020
Check for updates

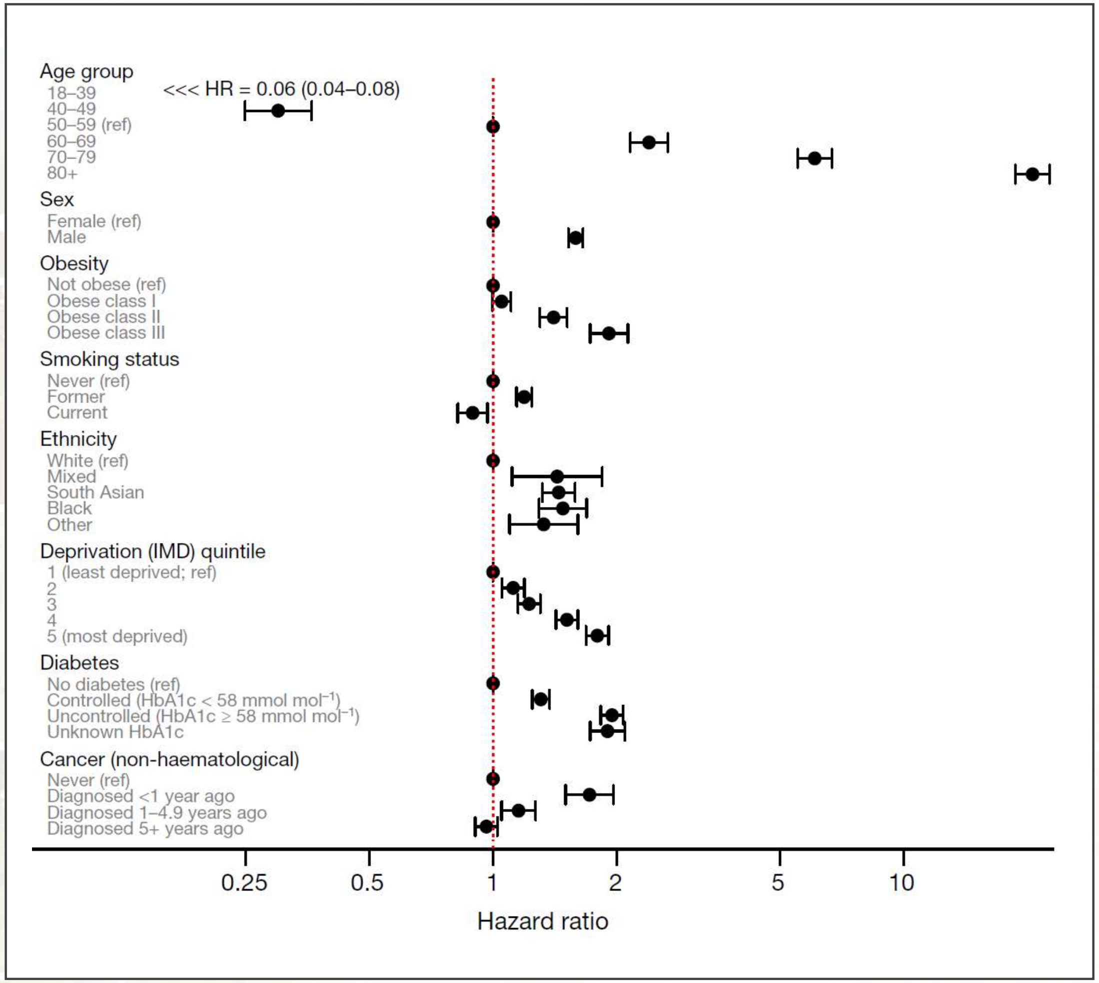
Elizabeth J. Williamson[1,6], Alex J. Walker[2,6], Krishnan Bhaskaran[1,6], Seb Bacon[2,6], Chris Bates[3,6], Caroline E. Morton[2], Helen J. Curtis[2], Amir Mehrkar[2], David Evans[2], Peter Inglesby[2], Jonathan Cockburn[3], Helen I. McDonald[1,4], Brian MacKenna[2], Laurie Tomlinson[1], Ian J. Douglas[1], Christopher T. Rentsch[1], Rohini Mathur[1], Angel Y. S. Wong[1], Richard Grieve[1], David Harrison[5], Harriet Forbes[1], Anna Schultze[1], Richard Croker[2], John Parry[3], Frank Hester[3], Sam Harper[3], Rafael Perera[2], Stephen J. W. Evans[1], Liam Smeeth[1,4,7] & Ben Goldacre[2,7] ✉

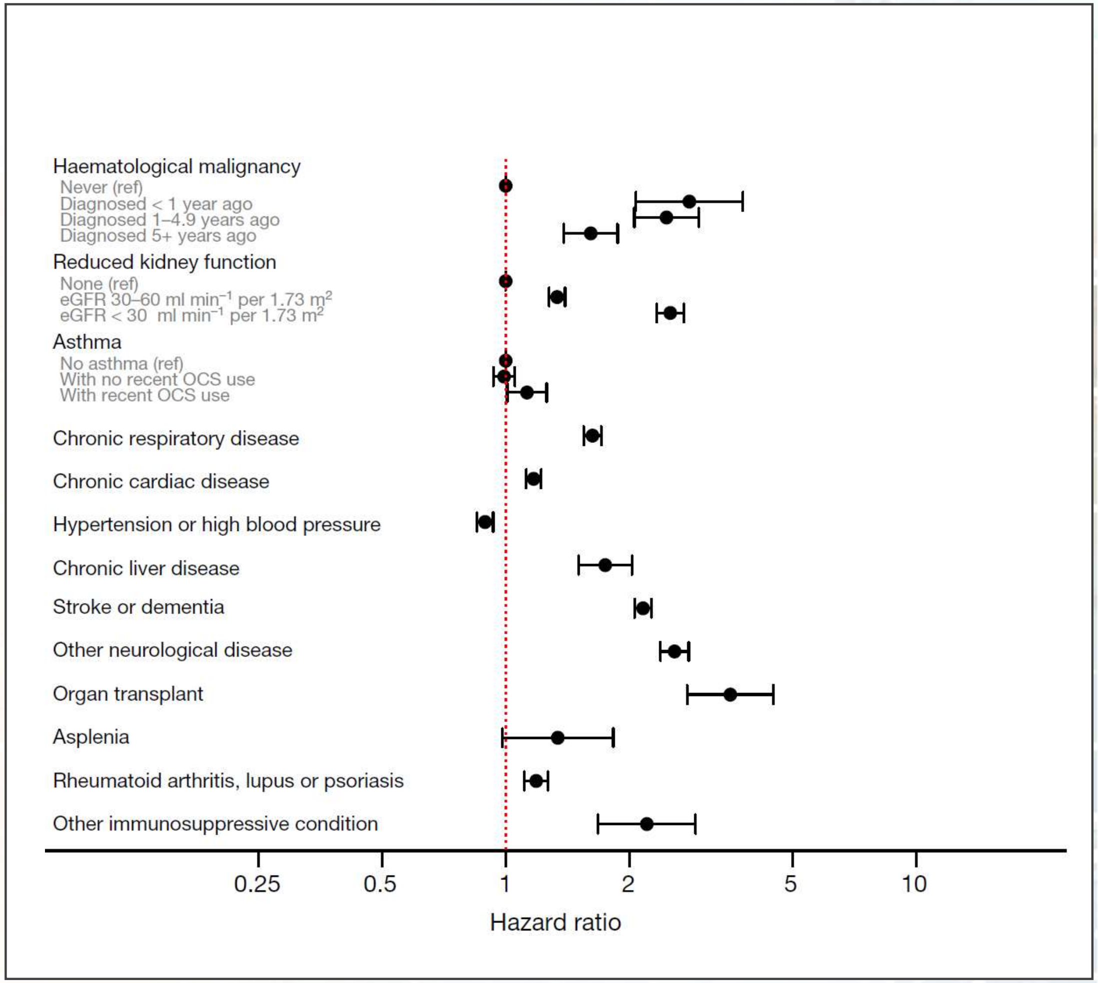
Coronavirus disease 2019 (COVID-19) has rapidly affected mortality worldwide[1]. There is unprecedented urgency to understand who is most at risk of severe outcomes, and this requires new approaches for the timely analysis of large datasets. Working on behalf of NHS England, we created OpenSAFELY—a secure health analytics platform that covers 40% of all patients in England and holds patient data within the existing data centre of a major vendor of primary care electronic health records. Here we used OpenSAFELY to examine factors associated with COVID-19-related death. Primary care records of 17,278,392 adults were pseudonymously linked to 10,926 COVID-19-related deaths. COVID-19-related death was associated with: being male (hazard ratio (HR) 1.59 (95% confidence interval 1.53–1.65)); greater age and deprivation (both with a strong gradient); diabetes; severe asthma; and various other medical conditions. Compared with people of white ethnicity, Black and South Asian people were at higher risk, even after adjustment for other factors (HR 1.48 (1.29–1.69) and 1.45 (1.32–1.58), respectively). We have quantified a range of clinical factors associated with COVID-19-related death in one of the largest cohort studies on this topic so far. More patient records are rapidly being added to OpenSAFELY, we will update and extend our results regularly.

Source: Williamson *et al* 2020 - *Nature*

- Age
- Sex
- BMI
- Smoking
- Ethnicity
- Area-level deprivation
- Blood pressure
- Hypertension
- Asthma
- Chronic heart disease
- Diabetes
- Cancer (blood)
- Cancer (non-blood)
- Kidney function
- Kidney dialysis
- Liver disease
- Stroke or dementia
- Oher neurological disease
- Organ transplant
- Asplenia
- Rheumatoid arthritis, lupus, or psoriasis
- Other immunosuppressive disease

OMITTED VARIABLE BIAS?





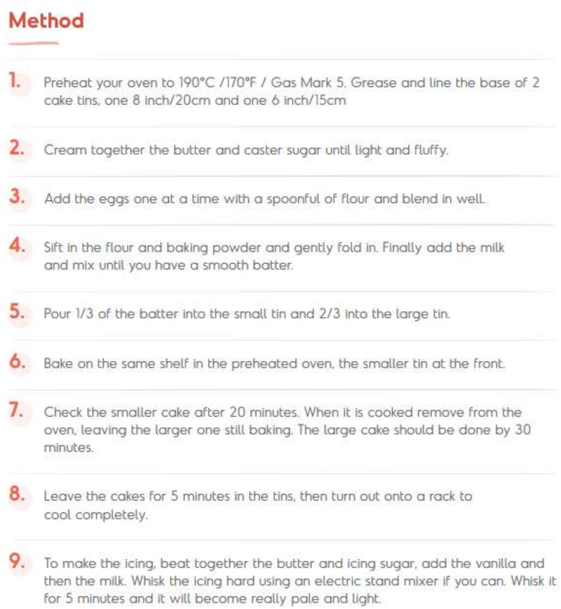
THE ESTIMAND INFORMS THE MODEL

- To estimate a causal effect, you must first ‘**identify**’ the causal effect **estimand** that your seek. The appropriate model is informed by that **estimand**!

ESTIMAND

E.g. The true difference in Y due to exposure

ESTIMATOR



Method

1. Preheat your oven to 190°C /170°F / Gas Mark 5. Grease and line the base of 2 cake tins, one 8 inch/20cm and one 6 inch/15cm
2. Cream together the butter and caster sugar until light and fluffy.
3. Add the eggs one at a time with a spoonful of flour and blend in well.
4. Sift in the flour and baking powder and gently fold in. Finally add the milk and mix until you have a smooth batter.
5. Pour 1/3 of the batter into the small tin and 2/3 into the large tin.
6. Bake on the same shelf in the preheated oven, the smaller tin at the front.
7. Check the smaller cake after 20 minutes. When it is cooked remove from the oven, leaving the larger one still baking. The large cake should be done by 30 minutes.
8. Leave the cakes for 5 minutes in the tins, then turn out onto a rack to cool completely.
9. To make the icing, beat together the butter and icing sugar, add the vanilla and then the milk. Whisk the icing hard using an electric stand mixer if you can. Whisk it for 5 minutes and it will become really pale and light.

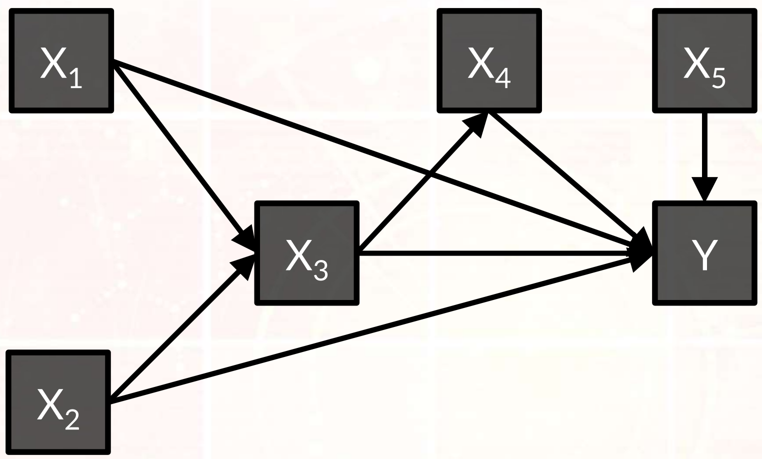
E.g. Your regression model

ESTIMATE

E.g. the estimated difference in Y from model coefficient

DIRECTED ACYCLIC GRAPHS

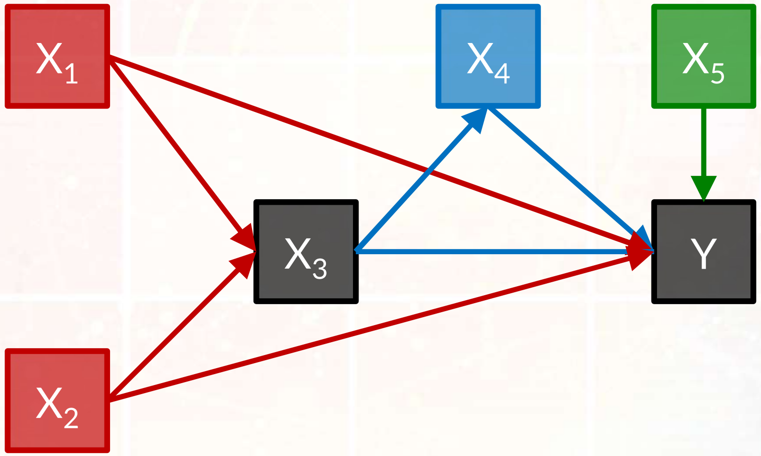
Directed Acyclic Graphs (DAGs) are nonparametric representations of the (hypothesised) causal relationships between variables



- Relationships between variables ('**nodes**') are represented by arrows ('**arcs**') creating **paths** between them
- Simple yet powerful way to encode **external knowledge** of the **data generating mechanism**

ESTIMATING CAUSAL EFFECTS

- To estimate the causal effect of $\mathbf{X_3}$ on **Y** (the '**focal relationship**'):
  - We want all **causal paths** open
  - And all **confounded paths** closed



ESTIMATING CAUSAL EFFECTS

- To estimate the causal effect of $\mathbf{X_3}$ on **Y** (the '**focal relationship**'):
  - We want all **causal paths** open
  - And all **confounded paths** closed
  - This means **conditioning** on all **confounders** but no **mediators**

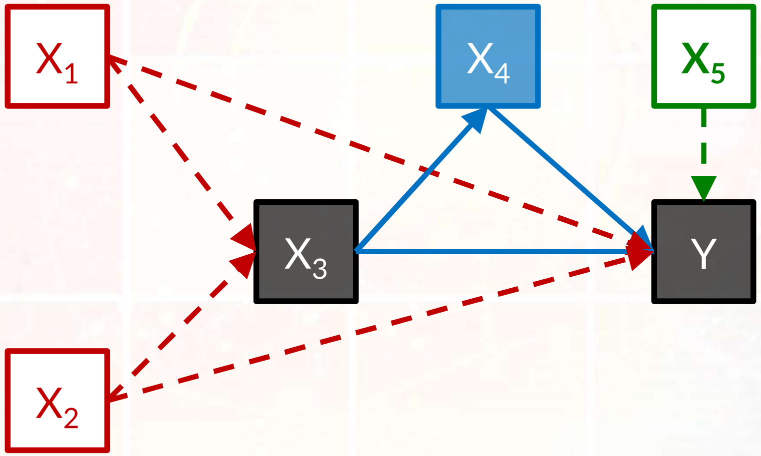
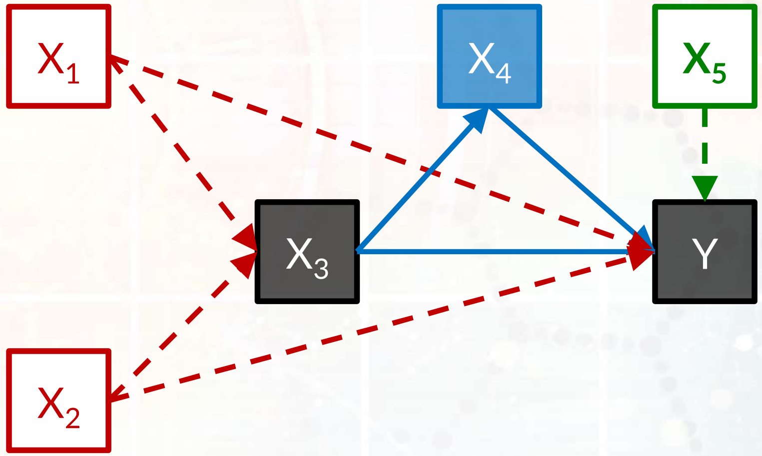


TABLE 2 FALLACY

Example: **Total causal effect** of **X_3** on **Y**:

- Model should include **confounders** (**X_1**, **X_2**) and **competing exposures** (**X_5**)



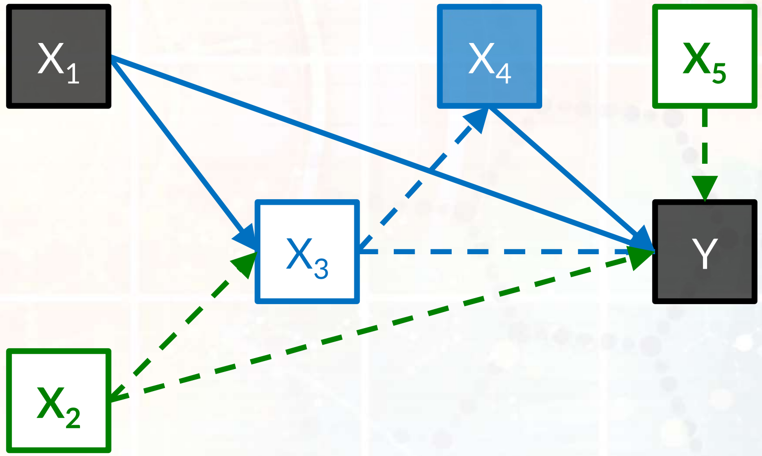
$$Y \sim X_3 + X_1 + X_2 + X_5$$

TABLE 2 FALLACY

Example: Total causal effect of X_3 on **Y**:

- Model should include **confounders** (X_1, X_2) and **competing exposures** (X_5)
- Would be **wrong** to ***interpret*** coefficients for other covariates ($\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_5}$), because they would require different adjustment sets!

E.g coefficient on $\mathbf{X_1}$ is **NOT** total causal effect of $\mathbf{X_1}$ on **Y**, due to conditioning on $\mathbf{X_3}$



$$\mathbf{Y \sim X_3 + X_1 + X_2 + X_5}$$

TABLE 2 FALLACY

- 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 has been dubbed the '**Table 2 Fallacy**'

American Journal of Epidemiology

Vol. 177, No. 4
DOI: 10.1093/aje/kws412
Advance Access publication:
January 30, 2013

Commentary

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.

Source: Westreich & Greenland 2013 - *Am J Epidemiol*

TABLE 2 FALLACY

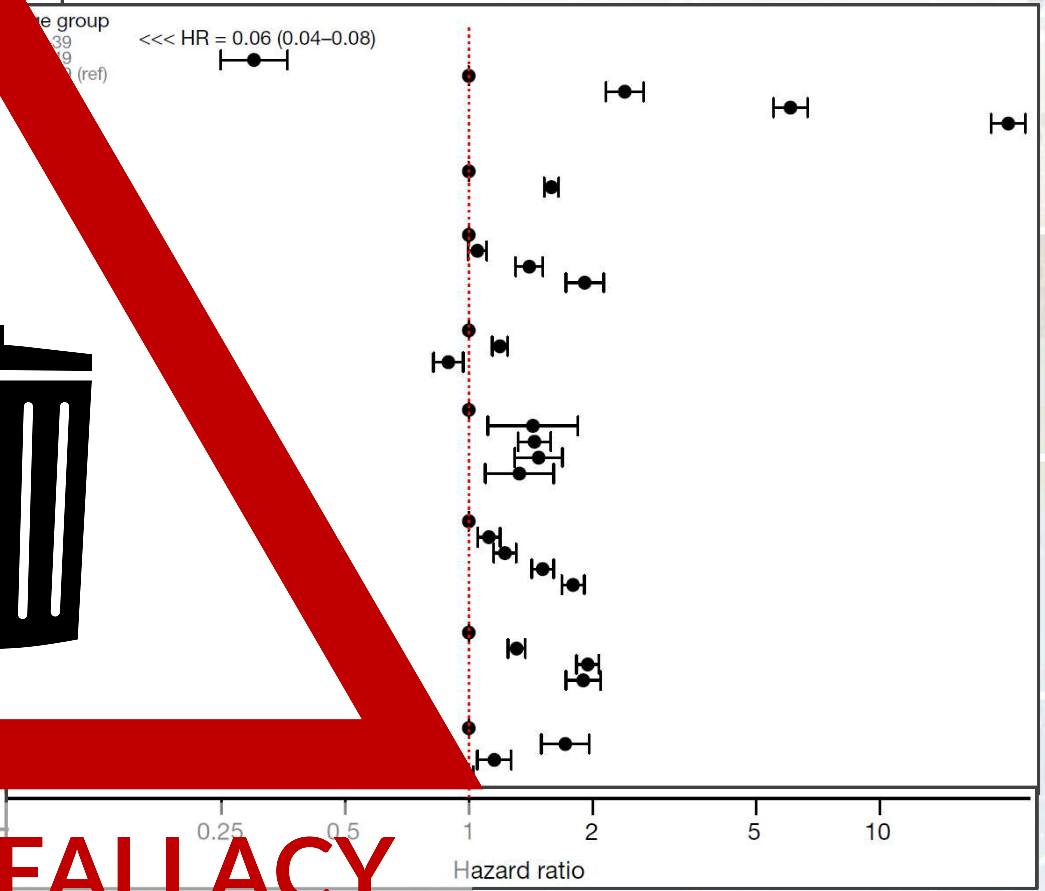
Article

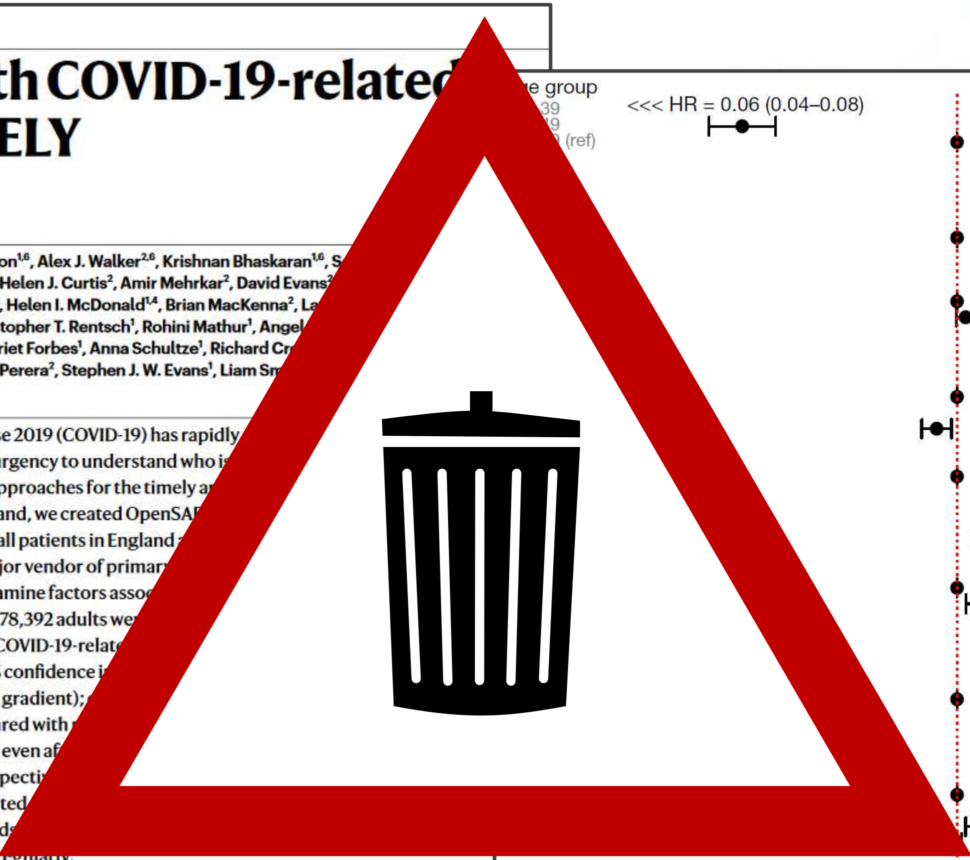
Factors associated with COVID-19-related death using OpenSAFELY

https://doi.org/10.1038/s41586-020-2521-4
Received: 15 May 2020
Accepted: 1 July 2020
Published online: 8 July 2020
Check for updates

Elizabeth J. Williamson[1,6], Alex J. Walker[2,6], Krishnan Bhaskaran[1,6], S
Caroline E. Morton[2], Helen J. Curtis[2], Amir Mehrkar[2], David Evans[2]
Jonathan Cockburn[3], Helen I. McDonald[1,4], Brian MacKenna[2], La
Ian J. Douglas[1], Christopher T. Rentsch[1], Rohini Mathur[1], Angel
David Harrison[5], Harriet Forbes[1], Anna Schultze[1], Richard Cr
Sam Harper[3], Rafael Perera[2], Stephen J. W. Evans[1], Liam Sm

Coronavirus disease 2019 (COVID-19) has rapidly
is unprecedented urgency to understand who i
this requires new approaches for the timely a
behalf of NHS England, we created OpenSAF
that covers 40% of all patients in England a
data centre of a major vendor of primar
OpenSAFELY to examine factors asso
care records of 17,278,392 adults we
19-related deaths. COVID-19-relat
ratio (HR) 1.59 (95% confidence i
(both with a strong gradient);
conditions. Compared with
were at higher risk, even af
1.45 (1.32–1.58), respecti
with COVID-19-related
More patient records
extend our results regularly.

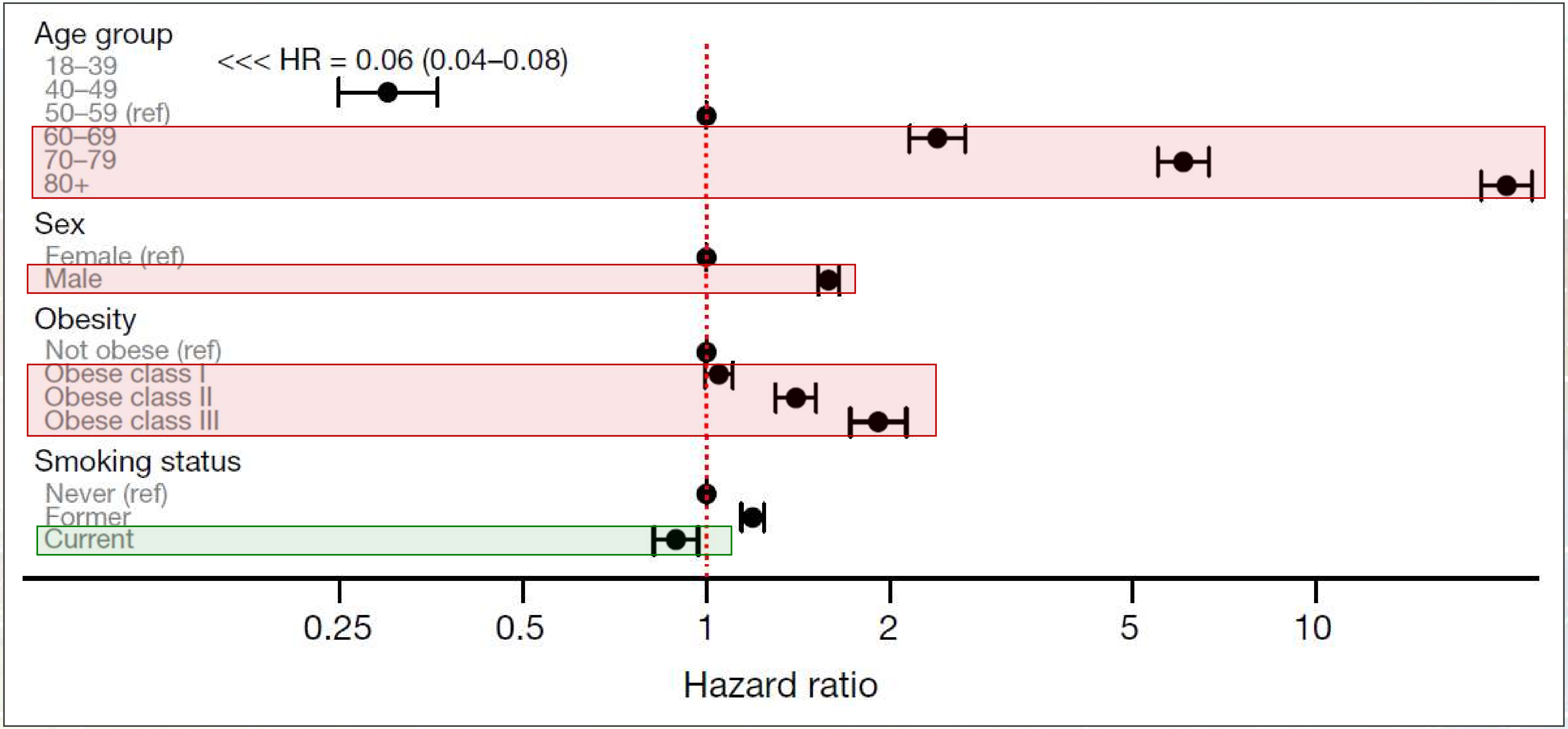




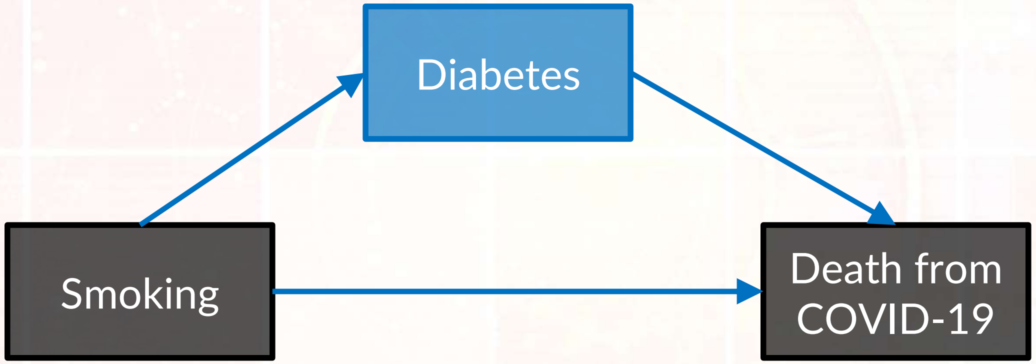
Source: Williamson *et al* 2020 - *Nature*

TABLE 2 FALLACY

TABLE 2 FALLACY

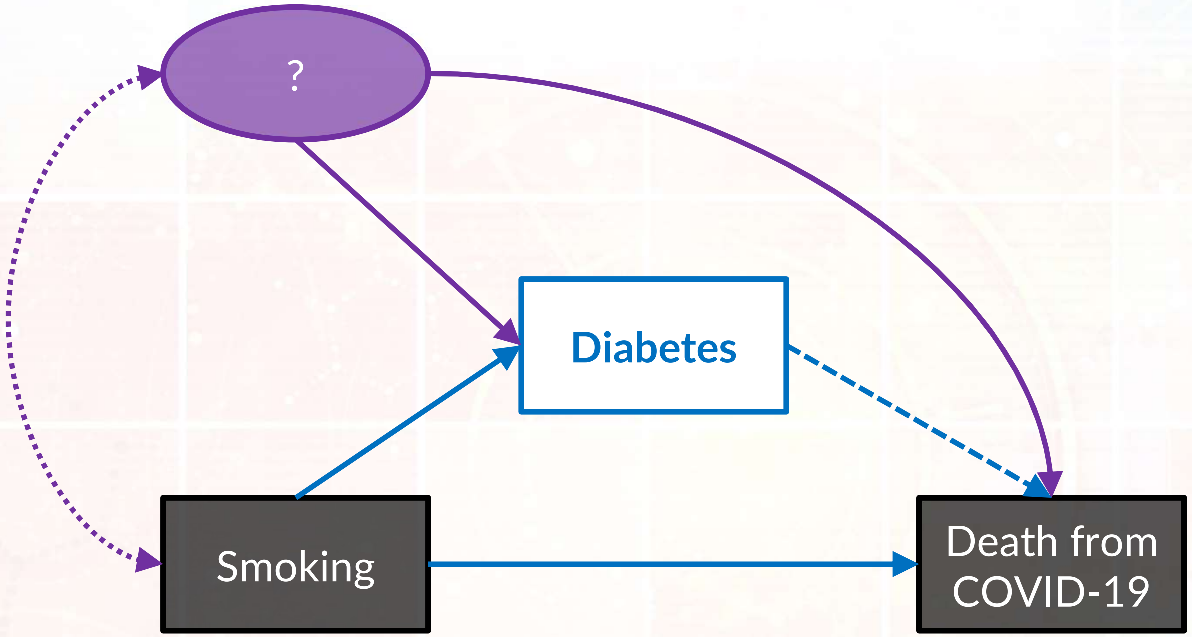


COLLIDER BIAS



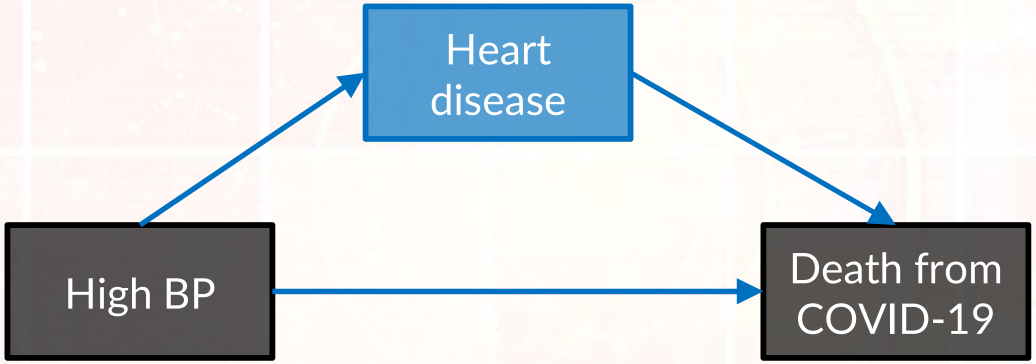
Smoking	Never	1.00 (ref)
	Former	1.43 (1.37–1.49)
	Current	1.14 (1.05–1.23)

COLLIDER BIAS



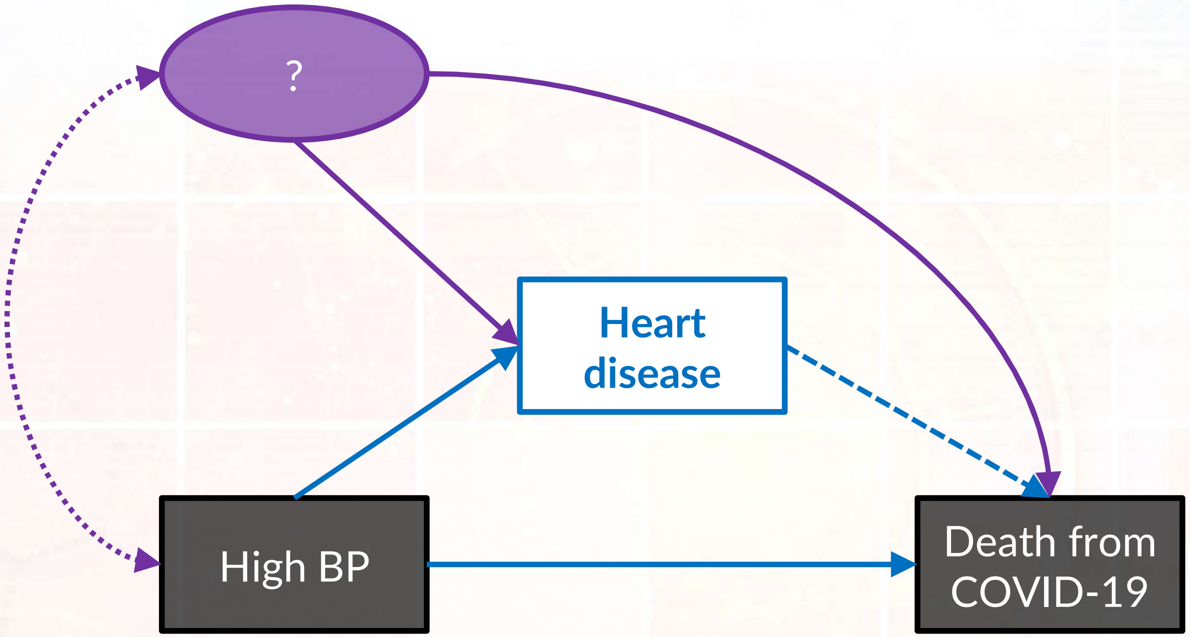
Smoking	Never	1.00 (ref)	1.00 (ref)
	Former	1.43 (1.37–1.49)	1.19 (1.14–1.24)
	Current	1.14 (1.05–1.23)	0.89 (0.82–0.97)

COLLIDER BIAS



Blood pressure	Normal	1.00 (ref)	1.00 (ref)
	High blood pressure or diagnosed hypertension	1.09 (1.05–1.14)	0.89 (0.85–0.93)

COLLIDER BIAS



Blood pressure	Normal	1.00 (ref)	1.00 (ref)
	High blood pressure or diagnosed hypertension	1.09 (1.05–1.14)	0.89 (0.85–0.93)

INAPPROPRIATE INTERPRETATION

Article

Factors associated with COVID-19-related death using OpenSAFELY

https://doi.org/10.1038/s41586-020-2521-4

Received: 15 May 2020

Accepted: 1 July 2020

Published online: 8 July 2020

Check for updates

Elizabeth J. Williamson[1,6], Alex J. Walker[2,6], Krishnan Bhaskaran[1,6], Seb Bacon[2,6], Chris Bates[3,6], Caroline E. Morton[2], Helen J. Curtis[2], Amir Mehrkar[2], David Evans[2], Peter Inglesby[2], Jonathan Cockburn[3], Helen I. McDonald[1,4], Brian MacKenna[2], Laurie Tomlinson[1], Ian J. Douglas[1], Christopher T. Rentsch[1], Rohini Mathur[1], Angel Y. S. Wong[1], Richard Grieve[1], David Harrison[5], Harriet Forbes[1], Anna Schultze[1], Richard Croker[2], John Parry[3], Frank Hester[3], Sam Harper[3], Rafael Perera[2], Stephen J. W. Evans[1], Liam Smeeth[1,4,7] & Ben Goldacre[2,7]✉

Coronavirus disease 2019 (COVID-19) has rapidly a
is unprecedented urgency to understand who is m
this requires new approaches for the timely analys
behalf of NHS England, we created OpenSAFELY—
that covers 40% of all patients in England and hold
data centre of a major vendor of primary care elec
OpenSAFELY to examine factors associated with C
care records of 17,278,392 adults were pseudonym
19-related deaths. COVID-19-related death was ass
ratio (HR) 1.59 (95% confidence interval 1.53–1.65)
(both with a strong gradient); diabetes; severe ast
conditions. Compared with people of white ethnic
were at higher risk, even after adjustment for othe
1.45 (1.32–1.58), respectively). We have quantified
with COVID-19-related death in one of the largest c
More patient records are rapidly being added to O
extend our results regularly.

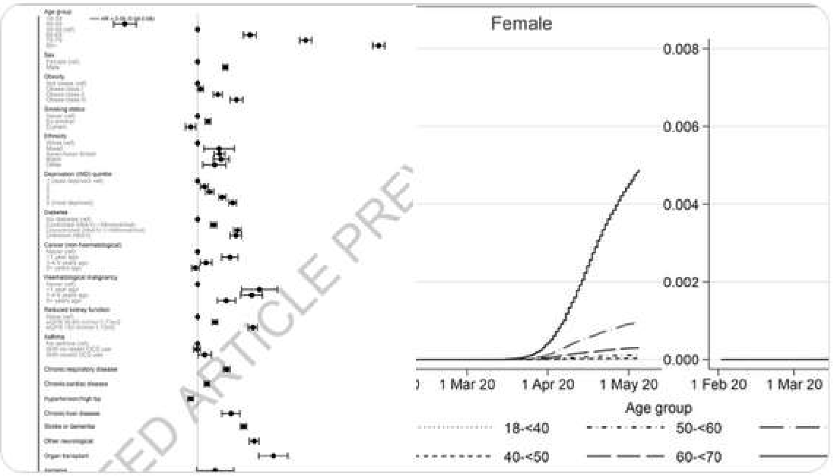
Our analyses provide a preliminary picture of how key demographic characteristics and a range of comorbidities—which were a priori selected as being of interest in COVID-19—are jointly associated with poor outcomes. These initial results may be used to inform the development of prognostic models. We caution against interpreting our estimates as causal effects. For example, the fully adjusted smoking hazard ratio does not capture the causal effect of smoking, owing to the inclusion of comorbidities that are likely to mediate any effect of smoking on COVID-19-related death (for example, chronic obstructive pulmonary disease). Our study has highlighted a need for carefully

Source: Williamson *et al* 2020 - *Nature*

PAVLOVIAN INFERENCE

Eric Topol @EricTopol

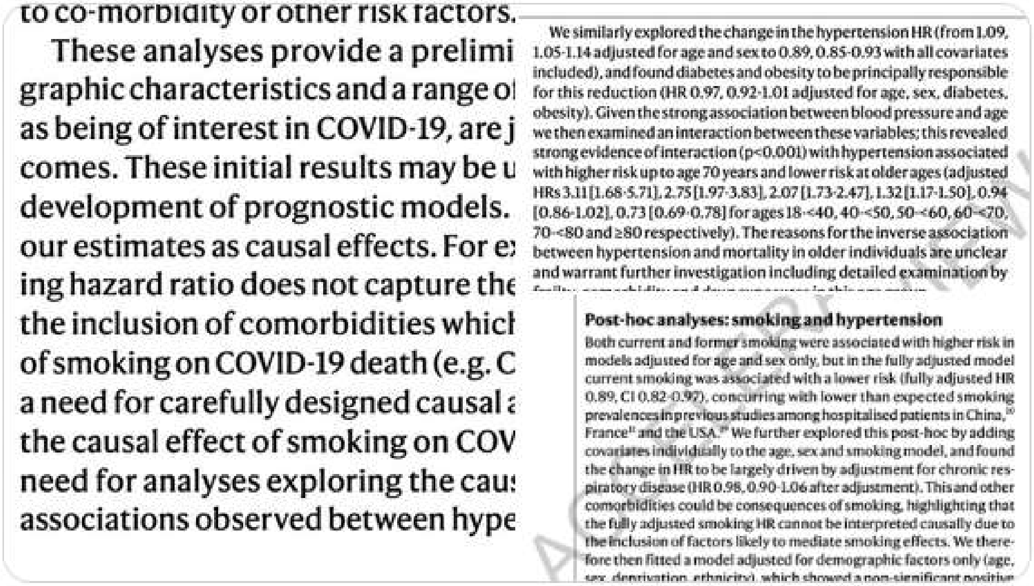
New @nature: the risk factors for dying from #COVID19 from >17 million people and ~11,000 deaths nature.com/articles/s4158... @bengoldacre and colleagues
importance of age, sex, race, diabetes, obesity, many other conditions
not risk factor: hypertension; current smoker protective



Ben CatchYourCough Goldacre @bengoldacre

Replying to @EricTopol and @nature

Thanks Eric, but we certainly don't think the data here show that smoking is "protective", as per the paper:



to co-morbidity or other risk factors.
These analyses provide a prelimi
graphic characteristics and a range o
as being of interest in COVID-19, are j
comes. These initial results may be u
development of prognostic models.
our estimates as causal effects. For e
ing hazard ratio does not capture the
the inclusion of comorbidities which
of smoking on COVID-19 death (e.g. C
a need for carefully designed causal a
the causal effect of smoking on COV
need for analyses exploring the caus
associations observed between hype

We similarly explored the change in the hypertension HR (from 1.09, 1.05-1.14 adjusted for age and sex to 0.89, 0.85-0.93 with all covariates included), and found diabetes and obesity to be principally responsible for this reduction (HR 0.97, 0.92-1.01 adjusted for age, sex, diabetes, obesity). Given the strong association between blood pressure and age we then examined an interaction between these variables; this revealed strong evidence of interaction (p<0.001) with hypertension associated with higher risk up to age 70 years and lower risk at older ages (adjusted HRs 3.11 [1.68-5.71], 2.75 [1.97-3.83], 2.07 [1.73-2.47], 1.32 [1.17-1.50], 0.94 [0.86-1.02], 0.73 [0.69-0.78] for ages 18-<40, 40-<50, 50-<60, 60-<70, 70-<80 and ≥80 respectively). The reasons for the inverse association between hypertension and mortality in older individuals are unclear and warrant further investigation including detailed examination by

Post-hoc analyses: smoking and hypertension

Both current and former smoking were associated with higher risk in models adjusted for age and sex only, but in the fully adjusted model current smoking was associated with a lower risk (fully adjusted HR 0.89, CI 0.82-0.97), concurring with lower than expected smoking prevalences in previous studies among hospitalised patients in China,[10] France[11] and the USA.[12] We further explored this post-hoc by adding covariates individually to the age, sex and smoking model, and found the change in HR to be largely driven by adjustment for chronic respiratory disease (HR 0.98, 0.90-1.06 after adjustment). This and other comorbidities could be consequences of smoking, highlighting that the fully adjusted smoking HR cannot be interpreted causally due to the inclusion of factors likely to mediate smoking effects. We therefore then fitted a model adjusted for demographic factors only (age,

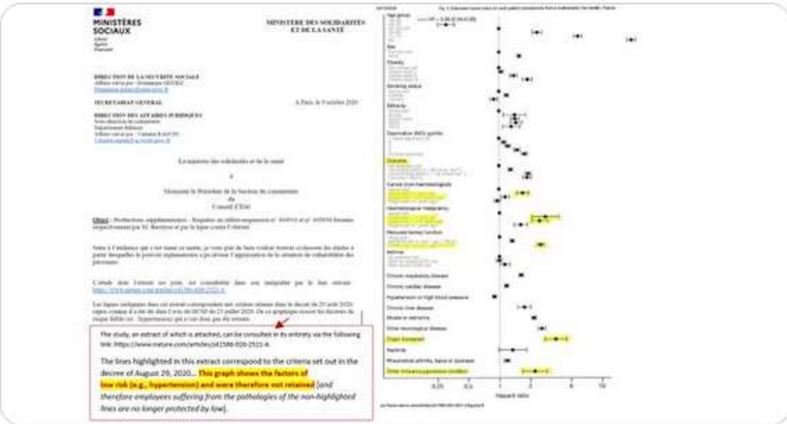
PAVLOVIAN INFERENCE

Olivier Berruyer
@OBerruyer

BIGGEST TABLE 2 FALLACY EVER

#EpiTwitter, do you remember the discussions around the @OpenSafely Williamson @nature study this summer?

The French @gouvernementFR admitted they used this study's Table 2 to withdraw protections from 100.000s of workers at risk of severe #COVID19!



Olivier Berruyer
@OBerruyer

Replying to @OBerruyer

VICTORY!

The French Supreme Court ruled in our favor, and overturned the government's decision, deeming it "neither coherent nor sufficiently justified" : all the Vulnerables are once again protected!

Thanks you all #EpiTwitter

And thanks to @bengoldacre for his answer: ⤵

Ben Goldacre <ben.goldacre@phc.ox.ac.uk> mer. 14 oct. à 10:44
À : Olivier Berruyer

Hi Olivier,

Our Nature paper provided one of the most detailed multivariate descriptions of relationships between individual-level factors and COVID outcomes, relatively early in the pandemic.

We cannot comment on the particular decision-making processes in France. In general, our recommendation would be that policy decisions should be based on the totality of evidence from different studies (and also, for example, take into account absolute levels of risk in different groups, rather than only relative risks/hazard ratios).

The "table 2 fallacy" occurs when inappropriate causal conclusions are drawn from a multivariable model. There were no such causal conclusions in our paper. However, in any case, this may not be a relevant consideration in applications that aim to simply identify groups at highest risk, for example to inform shielding policy. This is because even non-causal factors may be effective in identifying high-risk groups.

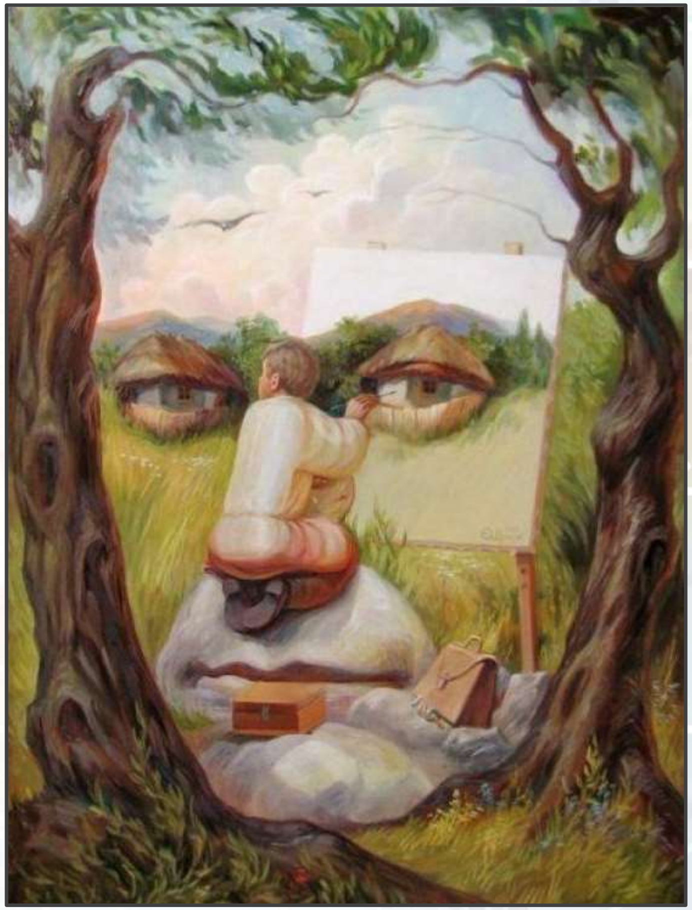
Best wishes,

Ben and OpenSAFELY team

Dr Ben Goldacre BA MA MSc MBBS MRCPsych MBE
Director, The DataLab
Fellow, Green Templeton College
Nuffield Dept of Primary Care Health Sciences
University of Oxford
Oxford OX2 6GG

TRANSPARENCY ≠ INTERPRETABILITY

- This example is from a *completely* **transparent** additive linear model
- But the same issues apply to all **predictive models**
- ‘**Features**’ are not what they appear to be!
  - They are context-specific **joint effects**, determined by the of variables inside *and outside* the model!
- This is one reason why **data-driven algorithms** are so sensitive to **contextual changes** and so vulnerable to **adversarial manipulation**



Artist: Oleg Shuplyak

SUMMARY

- **Data-driven algorithms** are excellent at identifying and utilizing *patterns* and *associations* within data
  - Excellent at encoding – and magnifying – **unfair** and **prejudiced** patterns
- Designing **explainable** and **interpretable** algorithms is one way to ensure more **reliable**, **ethical**, and **transportable** algorithms
- An **interpretable** algorithm is a **transparent** algorithm…
  - …But a **transparent** algorithm is **not necessarily interpretable**
- True **interpretability** requires **causal understanding**
  - Demonstrated by the **Table 2 Fallacy**
  - Coefficients are **transparent**, but may be **dangerously misleading**