Advancing Data Science to Examine the Causal Relationship Between Social Media Content and Adolescent Health Risk Behaviours

Digital Media for Behaviour Change Webinar Series

Dr Amrit Kaur Purba
Assistant Research Professor
University of Cambridge

What the existing evidence can - and can't - tell us

thebmj Research Education News & Views Campaigns Jobs

Social media use and health risk behaviours in young people: systematic review and meta-analysis

BMJ 2023 ; 383 doi: https://doi.org/10.1136/bmj-2022-073552 (Published 29 November 2023)
Cite this as: *BMJ* 2023;383:e073552

Article | Related content | Metrics | Responses | Peer review

Amrit Kaur Purba, PhD postdoctoral researcher, Rachel M Thomson, clinical research fellow, Paul M Henery, senior intelligence officer, Anna Pearce, senior research fellow, Marion Henderson, professor of child and youth wellbeing, S Vittal Katikireddi, professor of public health and health inequalities

No causal evidence

Poor exposure measurement

Don't know who is most vulnerable

Purba et al (2024) Nicotine and Tobacco Research
Purba et al (2024) European Journal of Public Health
Purba et al (2023) BMJ
Purba et al (2023) European Journal of Public Health

Understanding why evidence matters

Weak evidence base

Platforms delay action

Regulators lack clarity

Young people remain unprotected

Objective, granular social media data needed

Data donation

Data donation models

Data Download Packages (current)

Platform ⇨ Individual ⇨ Researcher

- Retrospective archive
- Manual download and upload
- High participant burden

API-based Data Donation (emerging)

Platform ⇨ Secure research environment

- Prospective and longitudinal
- Defined extract
- Low participant burden

What we know from psychology – and what public health adds

Psychology

- Essential for emotions, wellbeing, and lived experience
- Strong experimental tradition (RCTs) in other domains
- RCTs for social media: rarely feasible or ethical
- Cross-sectional, convenience samples → no population inferences
- Cannot quantify who is exposed to what or causal impact at scale

Why public health & epidemiology are essential for regulation

Public health & epidemiology

- Population-level: prevalence, risk ratios, vulnerability profiles
- Causal frameworks
- Equity: effects by age, sex, ethnicity, socioeconomic circumstance, algorithm settings etc
- COI-secure governance (commercial determinants model)
- Produces policy-ready evidence

Building the foundations for data donation

Feasibility and Acceptability of API-Based TikTok Data Donation Among UK Adolescents: A Mixed-Methods Simulation Study

Feasibility	Acceptability	Trust and ethics	Equity and digital literacy
Can young people actually complete API-based data donation?	Do they feel comfortable doing it?	What do they need to feel safe, respected, and in control?	Who might be excluded and why?

What we set out to learn

1. Does the simulation increase willingness to donate?
2. Who is more willing to donate before exposure to the simulation?
3. What aspects of the process shape willingness to donate?
4. What sharing models do youth prefer?
5. How do youth feel during the process?
6. How do we design equitable youth-centered consent?

Preliminary findings: youth and policy perspectives

Youth Advisory Groups

GenerationR
young people improving Research

YOUNG DYNAMOS

- API-style flow = easy and usable
- Comfort moderate-high; trust conditional
- TikTok less personal; Instagram more sensitive
- Need simple, clear explanations (what/why/how)
- Stronger altruism

Policy Advisory Group

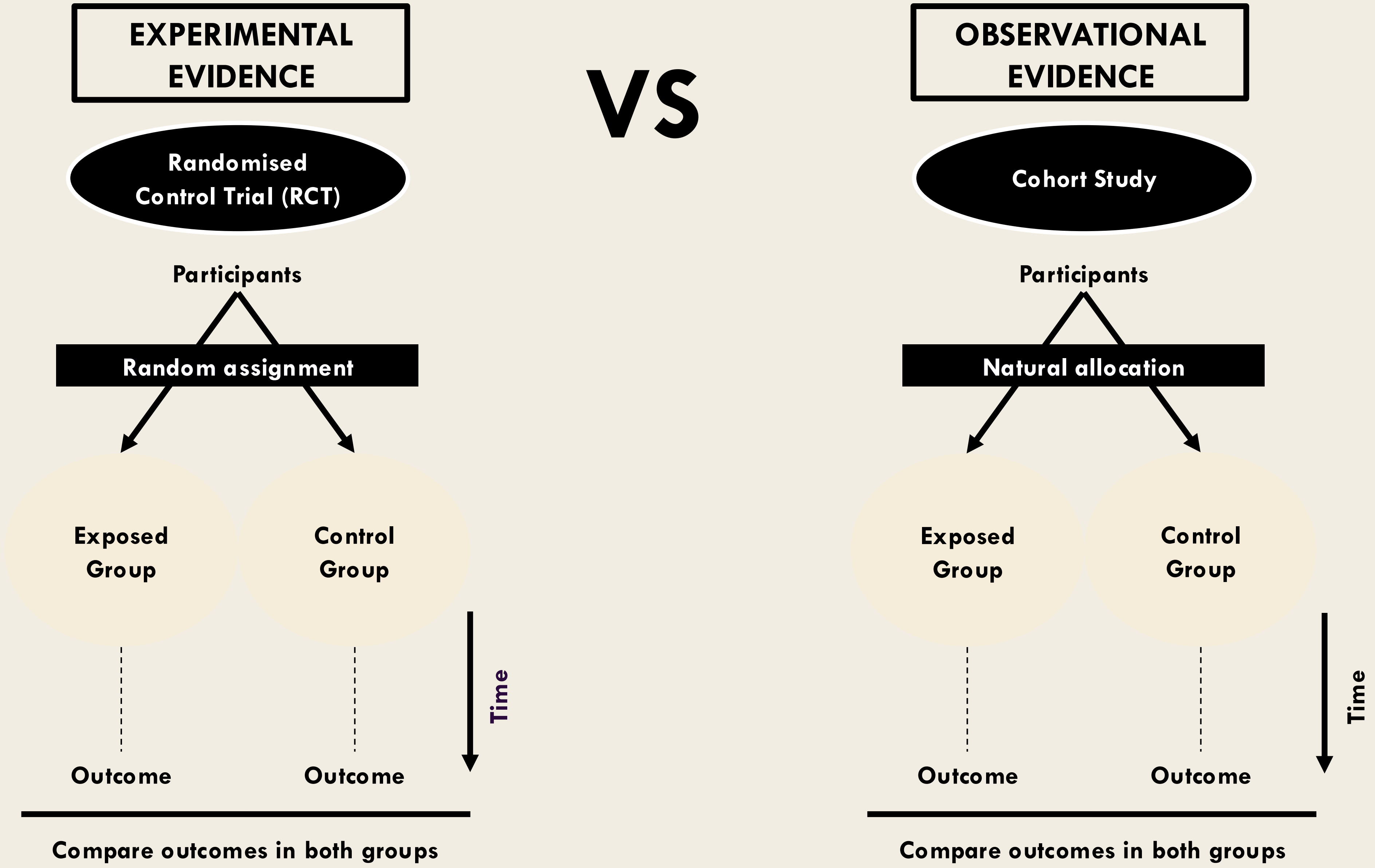
Office for Health Improvement & Disparities
Department of Health & Social Care
Department for Education
Department for Science, Innovation & Technology
NHS Greater Glasgow and Clyde
NHS England
Centre for Protecting Women Online
London School of Hygiene & Tropical Medicine
ADHD UK
Alliance 4 Children
Ofcom

- Strong support for objective exposure data
- Priorities: clarity, transparency, governance
- Need age-appropriate explanations
- Equity concerns in recruitment
- Safeguarding: signposting + support

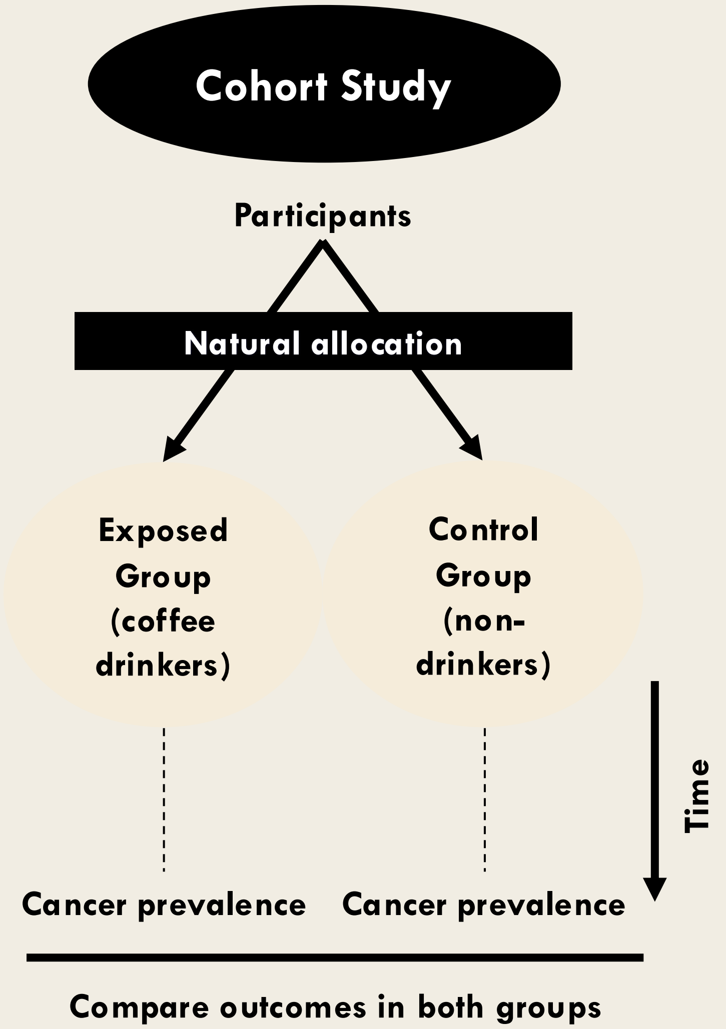
What this study delivers

- Evidence on **feasibility + acceptability** of API-based data donation
- Insights into **where youth hesitate** & what they need to feel safe
- Identification of **privacy misunderstandings**
- Understanding of **equity risks** & inclusion challenges
- Guidance for **trust-building, clarity & transparency**
- Direct input into **consent design & user journey**
- Foundations for **ethical, scalable national deployment**

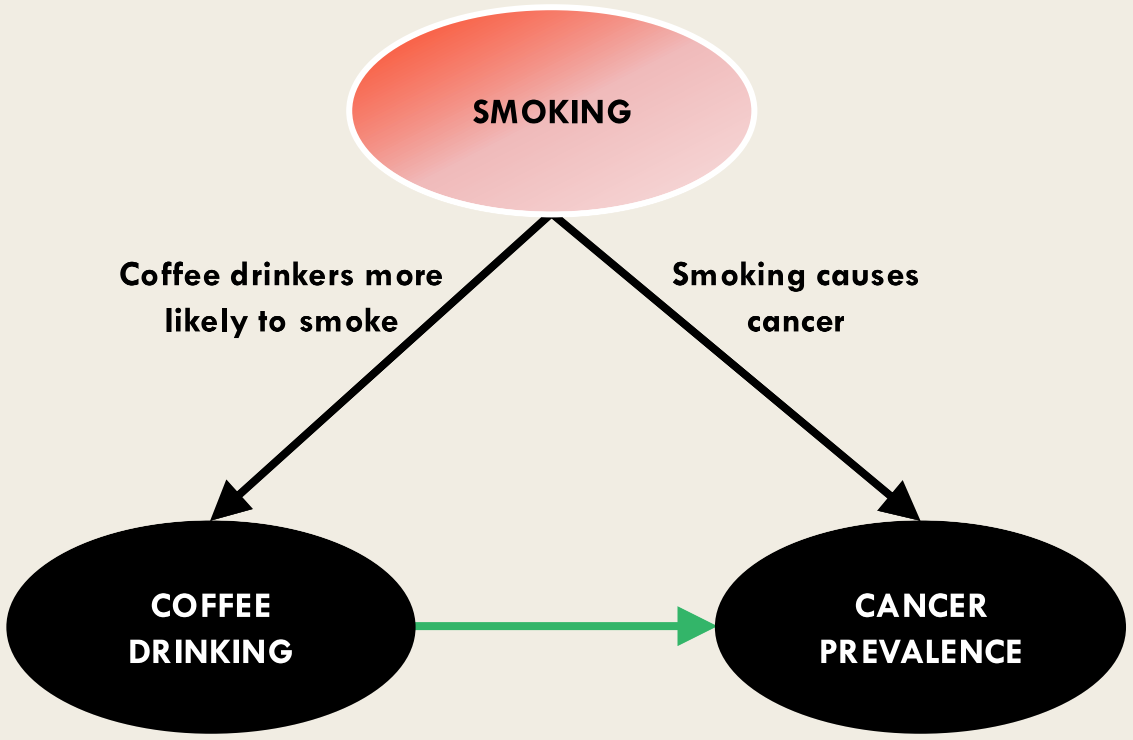
Answering causal questions



Why observational data can mislead



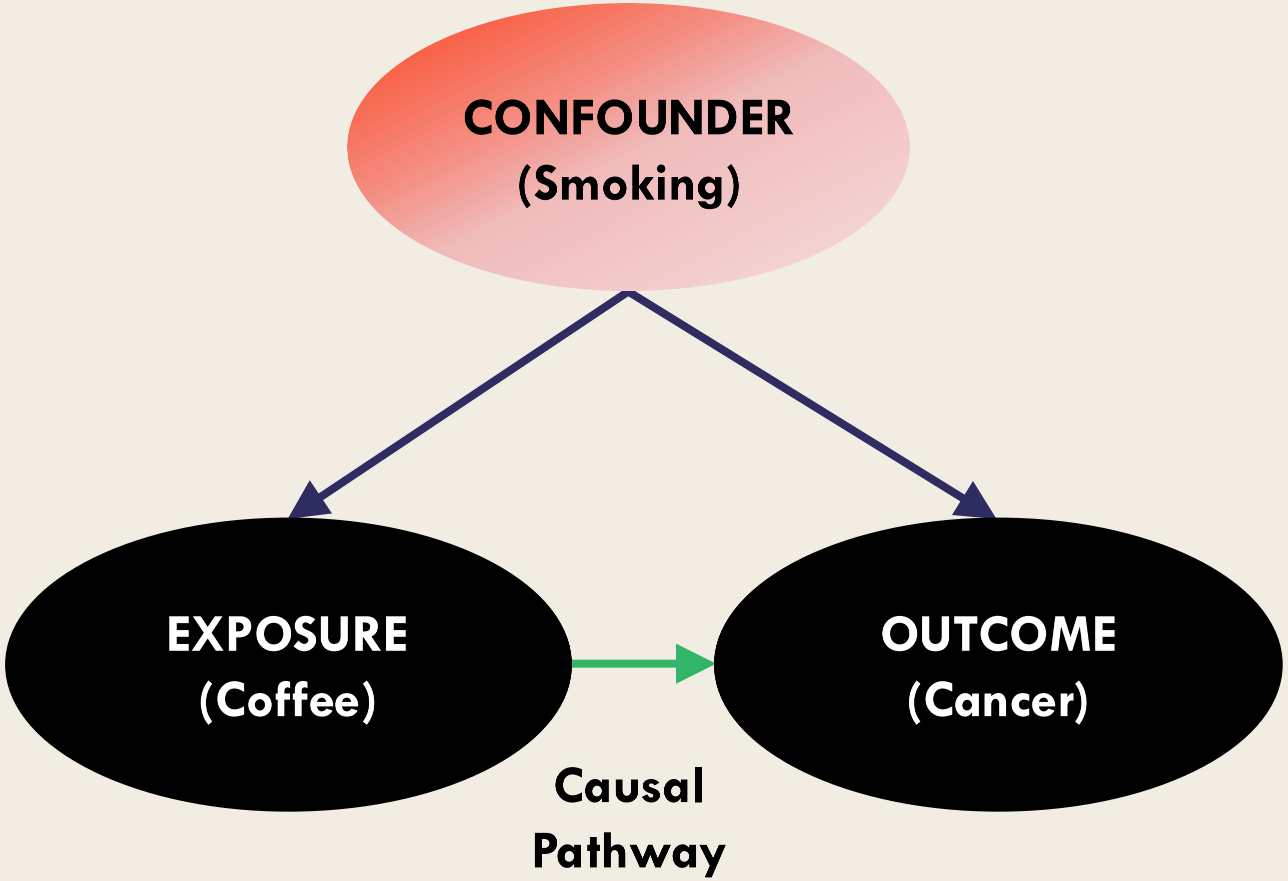
Result - Prevalence of cancer in coffee drinkers is higher than in non-drinkers…. really?



Why randomisation works and what we do without it

EXPERIMENTAL EVIDENCE		OBSERVATIONAL EVIDENCE
Randomised Control Trial (RCT)	VS	Cohort Study
Participants		Participants
Random assignment		Natural allocation
⇩		⇩
EXCHANGABLE 'COMPARABLE' GROUPS		**CONDITIONALLY EXCHANGABLE 'COMPARABLE' GROUPS**

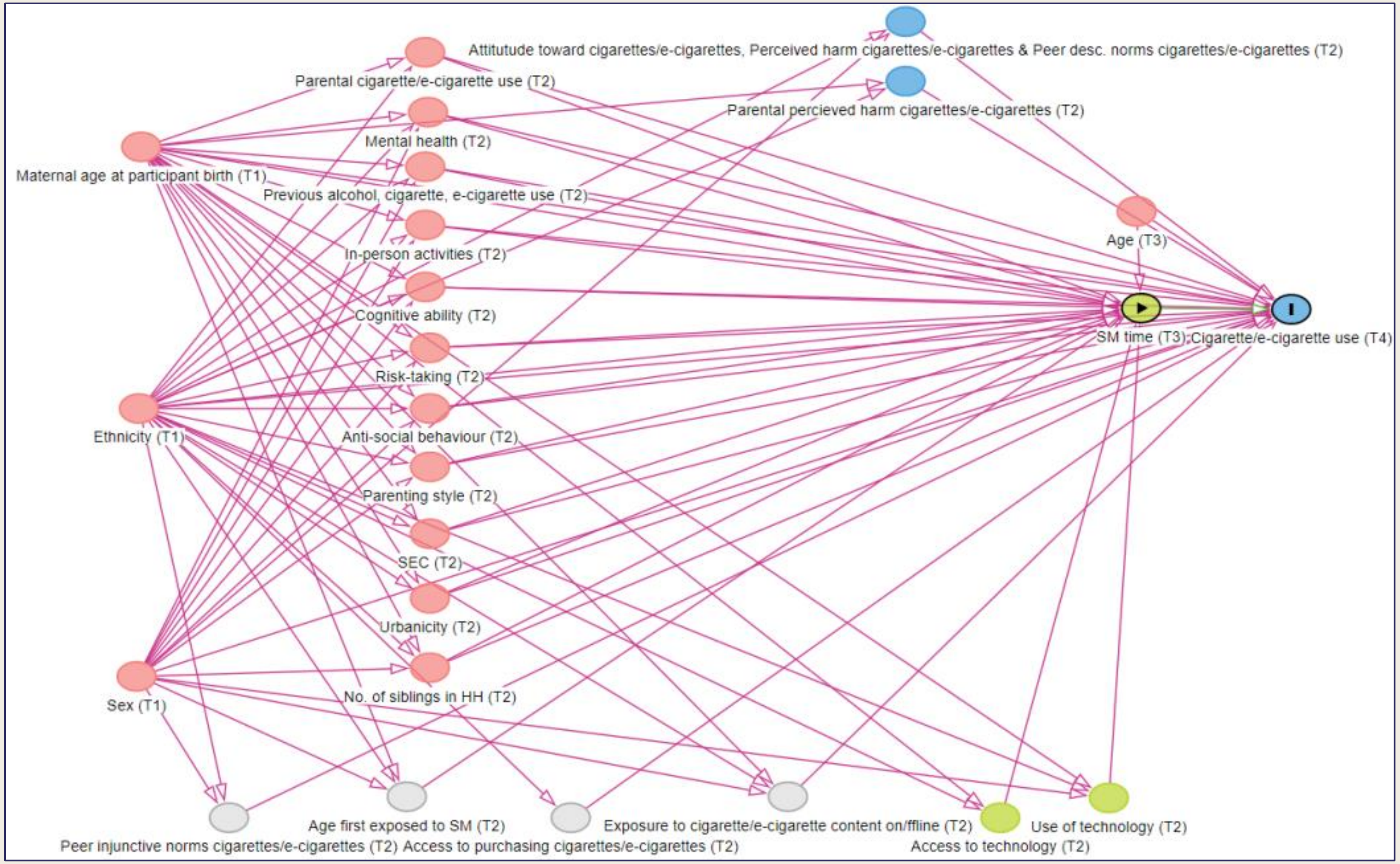
What is a confounder?



1. Identify the confounders
2. Measure them accurately
3. Adjust for them in the analysis

Directed Acyclic Graphs

DAG illustrating the hypothesised relationship between social media use at 14 years and alcohol use at 17 years



Why not just do an RCT?

Unethical to expose youth to harmful content

Personalised algorithms prevent true randomisation

Recruitment biases limit representativeness

Low-base-rate harms require population scale data

Trials move too slowly for fast-changing platforms

Findings may not reflect real-world digital environments

What is a target trial?

Target trial emulation

Design the RCT we *wish* we could run ⇨ Apply it to real-world data

Target Trials mimic the design of an RCT

DAGs give us valid confounder control

Cohorts provides population level data

Data donation gives us exposure detail

⇩

Causal evidence

From target trial design to causal effect estimates

Target trial

Target trial specified

- Eligibility
- Exposure & comparator
- Time zero & follow-up
- Outcome
- Estimand: ATE (>CATE)

Estimation strategy

- Weighting-based
- Outcome-model
- Doubly robust
- Extensions (fixed effects, sensitivity analyses)

Assumptions & outputs

Causal assumptions

- Consistency
- Positivity
- Conditional exchangeability

Outputs

- Interpretable causal effect estimates

Design first. Estimation second

DIGITAL DETERMINANTS OF HEALTH HUB

RESEARCH | POLICY | PRACTICE

Does exposure to alcohol, drug use, and antisocial behaviour content on TikTok *actually* cause changes in young people's real-world behaviour?

Delivering causal evidence

James Lind Alliance Priority-Setting exercise

Evidence review

Longitudinal health + behavioural data

13-16 y (n = 300)   18-24 y (n = 2,000)

TikTok data donation

+

AI + human labelling to detect risk behaviour content

Real-time questions on risk behaviour content exposure

Causal inference

Youth Advisory Group

Policy Advisory Group

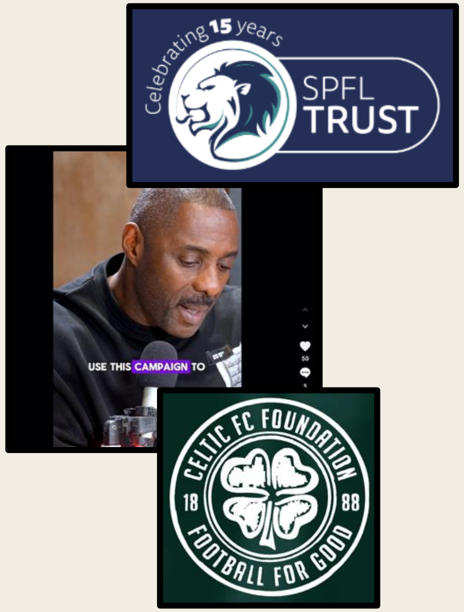
From insight to impact

Policy workshop

School visits +
Youth sessions

Public engagement
campaign

Closing public
webinar



Thank you! Any Questions?

Email: amritkaur.purba@mrc-cbu.cam.ac.uk

linkedin.com/in/amrit-kaur-purba-/

amritkaurpurba.com

Digital Determinants of Health Hub LinkedIn

wellcome

the HUO FAMILY FOUNDATION

UKRI Medical Research Council

UNIVERSITY OF CAMBRIDGE