Causal Methods at LSHTM Department of Medical Statistics

Bianca De Stavola, Rhian Daniel and Richard Silverwood

CTU-MSD Methodology Workshop

29 January 2013

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

NCRM National Centre for Research Methods

PATHWAYS biosocial influences on health

Improving health worldwide

www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Contents

- PATHWAYS
- Other causal work

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Introduction

- ESRC National Centre for Research Methods (NCRM) node
- Funded until October 2014.
- Aims:
    - Identify the pathways that link socio-demographic circumstances and biological disadvantage to health
    - Develop and disseminate methods for the investigation of pathways between social and health related processes
    - Offer training for social scientists in the use of biomedical data to maximise returns on new data investments
- http://pathways.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

People

- Professor Emily Grundy (Director)
- Professor Bianca De Stavola (Deputy Director)
- Professor Mike Kenward (Co-Director)
- Dr George Ploubidis (Co-Director)
- Dr Sanna Read
- Dr Richard Silverwood
- Ms Rohini Mathur (PhD Student)
- Dr Rhian Daniel (Collaborator)
- Dr Frank Dudbridge (Collaborator)
- And further collaborators...

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Substantive applications

1. To what extent does stress, social support and health related behaviour mediate the effect of fertility history and childhood circumstances on later life health?
2. To what extent does marital history mediate the association between childhood and early life circumstances and health in mid life?
3. Social disadvantage and infant mortality: effect modification by birthweight or selection bias?
4. Is alcohol use causally related to fibrinogen level?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Substantive applications

5. Collaboration with LEMMA (Longitudinal Effects, Multilevel Modelling and Applications)
    - ‘E-books for causal modelling and missing data methods’
    - Using their Stat-JR software environment to developing an ebook version of our practical on using the g-computation formula to investigate mediation.
6. Collaboration with PEPA (Programme Evaluation for Policy Analysis)
    - ‘Do income and wealth mediate associations between fertility histories and later life health?’
    - Extension of Project 1.
7. Collaboration with the Scottish Longitudinal Study
    - ‘From birth to childhood: investigating socio-economic differences in health trajectories via administrative data’
    - Extension of Project 3.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Methodological challenges

a) Complex structures
b) Measurement error
c) Missing data
d) Unmeasured confounding

Methods being explored

- **Structural equation models** and **g-computation of parametric causal models** can deal with a)-c), but impose strong modelling assumptions which should be explored through sensitivity analyses.
- Alternative models such as **marginal structural models** and **structural nested models** can be fitted semi-parametrically, relaxing some of these assumptions.
- They can also relax some of the unmeasured confounding assumptions implicit in SEMs.
- **Instrumental variable** based methods can deal with d) but require appropriate instruments.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- Light-to-moderate alcohol consumption thought to be protective for CVD based on observational studies, but this could be due to confounding or reverse causation.
- Collaboration of ~50 studies including ~200,000 individuals with data on an allele in the ADH1B gene, CVD biomarkers, diabetes, stroke and CHD.
- Conventional instrumental variable/Mendelian randomisation approach only allows investigation of linear associations.
- Developed a novel method using a local average treatment effect (LATE) that permits the investigation of nonlinear associations.
- Estimates causal effects for a series of discrete levels of alcohol consumption and tests whether these alcohol effects (derived from the genetic instrument) followed a linear or quadratic form.
- Requires the (strong) assumption that the gene has the same effect on alcohol consumption in all individuals.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Courses

All available in ‘Introductory’ and ‘Advanced’ flavours:

- Biomarkers (1 day)
- Genetics (1 day)
- Concepts and Methods in Causal Inference (1 or 2 days)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Other causal work

Bianca and Rhian

- gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula
- Using causal diagrams to guide analysis in missing data problems
- Methods for dealing with time-dependent confounding
- Causal mediation analysis with multiple causally-ordered mediators
- Mediation by structural equation modelling or causal inference: What is the difference?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

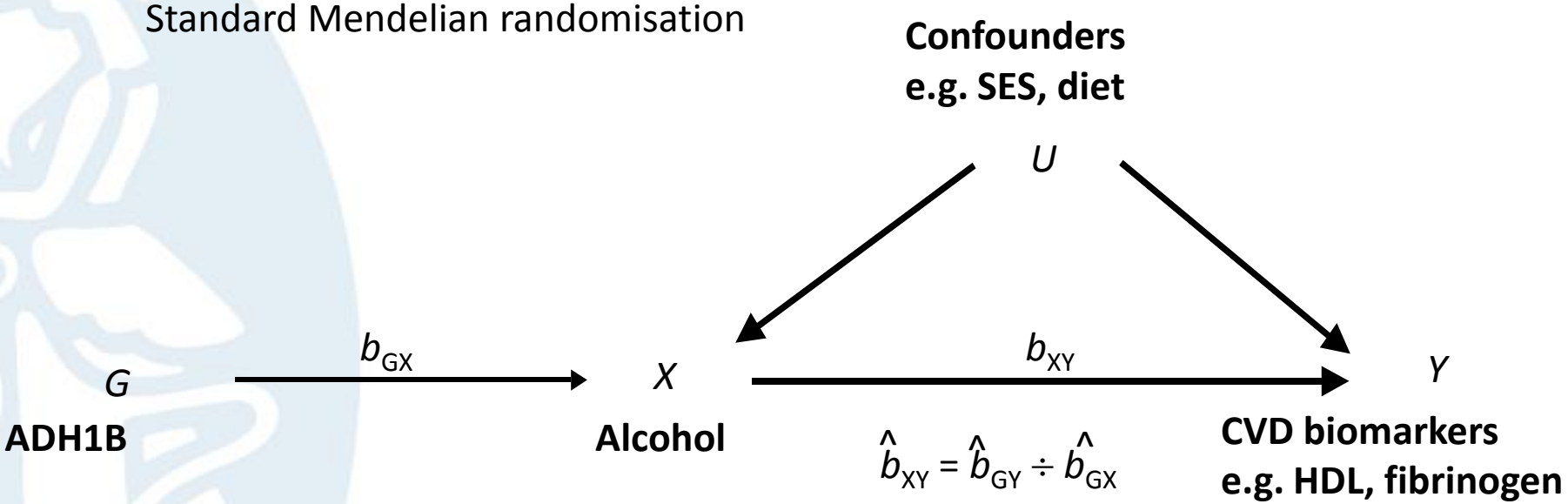
Extra slides

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

Standard Mendelian randomisation



Assumptions:

1. The genotype of *ADH1B* is associated with exposure of interest, alcohol
2. G is independent of the confounding factors (that confound alcohol - *biomarker* association)
3. ADH1B is related to the outcome *only* via its association with alcohol

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- Assume that for each subject, Y(X=x) is the outcome (SBP) that would be observed by setting X (alcohol) to the value x.
- Average treatment effect is

$$E\left[Y_i(x+1) - Y_i(x)\right]$$

  Meaningful when this is constant over the range of X.

- Define local average treatment effect (LATE) for exposure level j as

$$\tau_j = E\left[Y_i(j) - Y_i(j-1) \mid X_i(1) \geq j > X_i(0)\right]$$

- Need to identify individuals whose X (alcohol) is changed from 'at least j' to 'less than j' by changing their genotype.
- Then what change in Y (SBP) is expected by changing their X (alcohol) from j-1 to j?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- Assume a linear model relating X to G

$$X = \beta_{GX} G + \varepsilon_X$$

- Then discretise X into units of β_{GX}

$$X' = \left\lfloor \frac{X}{\beta_{GX}} \right\rfloor$$

- Changing G always changes X' by exactly one unit: everyone becomes a 'complier'.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- After discretising, for each *j* select the individuals with

  X' = j - 1 and G = 0

  or

  X' = j and G = 1

- By construction, these individuals are precisely those with $X(1) \geq j > X(0)$.
- Assume a (local) linear model for Y on X'

$$Y = \beta_{XY} X' + \varepsilon_Y$$

- Then for individuals with X' = j – 1 and G = 0 we know Y(j - 1) = Y and Y(j) = Y + β_{XY}
- And for individuals with X' = j and G = 1 we know Y(j) = Y and Y(j - 1) = Y - β_{XY}
- So linear regression of Y on X' yields the LATE for X' = j

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- We get a LATE, with its standard error, for each 'bin' *j* of X
- These could be combined by meta-analysis to get an average LATE, comparable to the overall average treatment effect
- General test of heterogeneity (e.g. Cochran Q) can detect a nonlinear causal effect
- Meta-regression of LATE on *j* can detect a linear trend in the LATEs
- As each LATE is itself the slope in a linear model, a linear trend suggests an overall quadratic shape
- N.B. Weak instrument bias in each LATE.
- N.B. We must estimate the bin size β_{GX} from the same data ($\longrightarrow$ bootstrap)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

Additional assumptions:

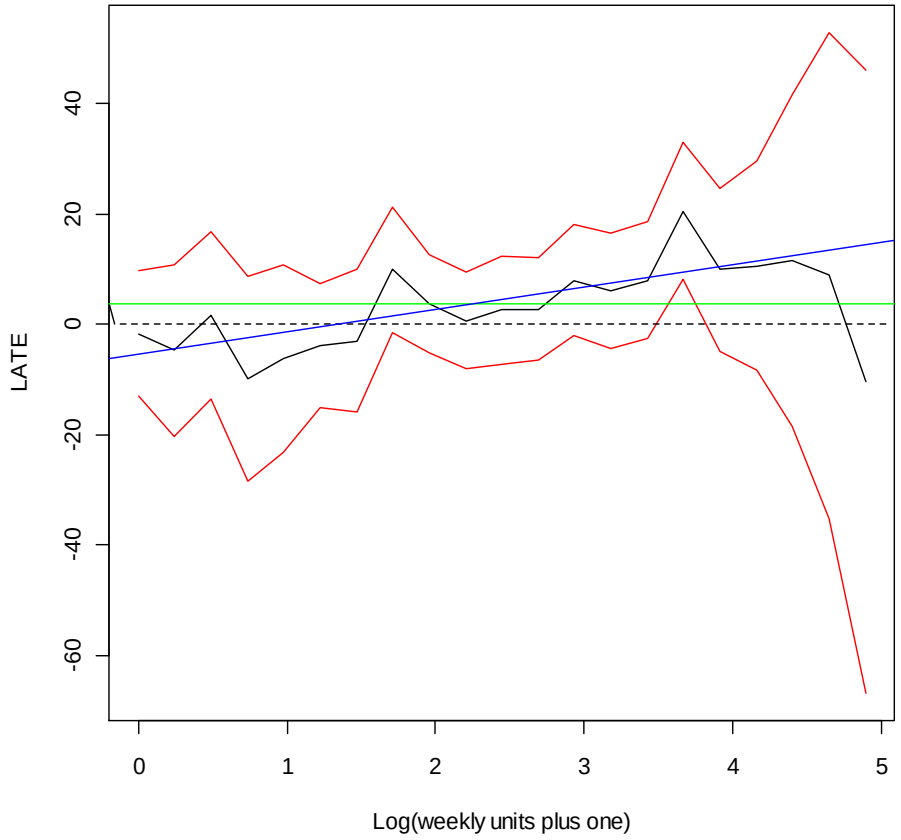
- Effect of instrument G (genotype) on phenotype X (alcohol) is the same for all subjects
    - No individual-level effect modifiers
    - Homogeneity across studies contributing to meta-analysis
    - Homogeneity of effect across the range of X – supported by observational data
- Monotonicity – the gene always changes alcohol consumption in the same direction

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

Results: SBP



Meta-analysis:	Mean LATE	3.3 (95% CI 0.7, 6.7)
Meta regression:	LATE intercept	-7.0 (95% CI -12.6, 1.1)
	LATE slope	4.6 (1.7, 6.9)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

PATHWAYS

Testing for a non-linear causal effect in a Mendelian randomisation study

- We have also conducted a lot of simulations.
- The method appears to work well under its assumptions.
- Under quite strong violations of the homogeneity assumption it appears to give valid tests of no quadratic effect (little bias but some under-coverage of its CI).

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE