CAUSAL INFERENCE FROM OBSERVATIONAL STUDY DESIGNS – AND APPLICATIONS TO STUDIES OF ALCOHOL CONSUMPTION AND SEXUAL RISK

Suzi Gage

University of Liverpool

UNIVERSITY OF LIVERPOOL

Structure of the talk

- How to get at causality in complex associations
- Methods to strengthen causal inference from observational data
  - Mendelian randomization
  - Negative/positive controls
  - Cross contextual studies
  - Sibling designs
  - Natural experiments
- Discussion

Background

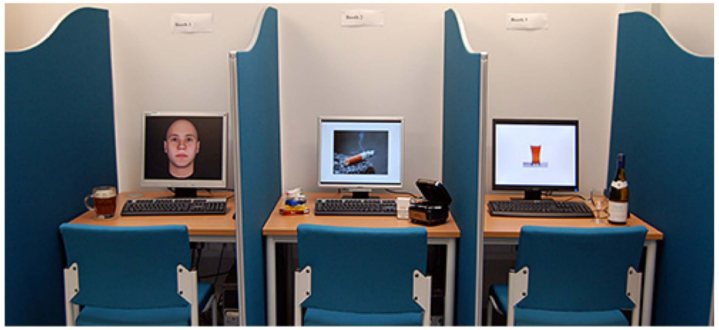

Levels of Evidence Pyramid
Systematic Reviews
Randomized Controlled Trials
Cohort Studies
Case-Control Studies
Case Series, Case Reports
Editorials, Expert Opinion

How we ascertain causality

- Experimental studies
  - Often only proxy outcomes
- Animal studies
  - Transferability
- Observational studies
  - Residual confounding

Triangulation

Observational studies



Experimental studies

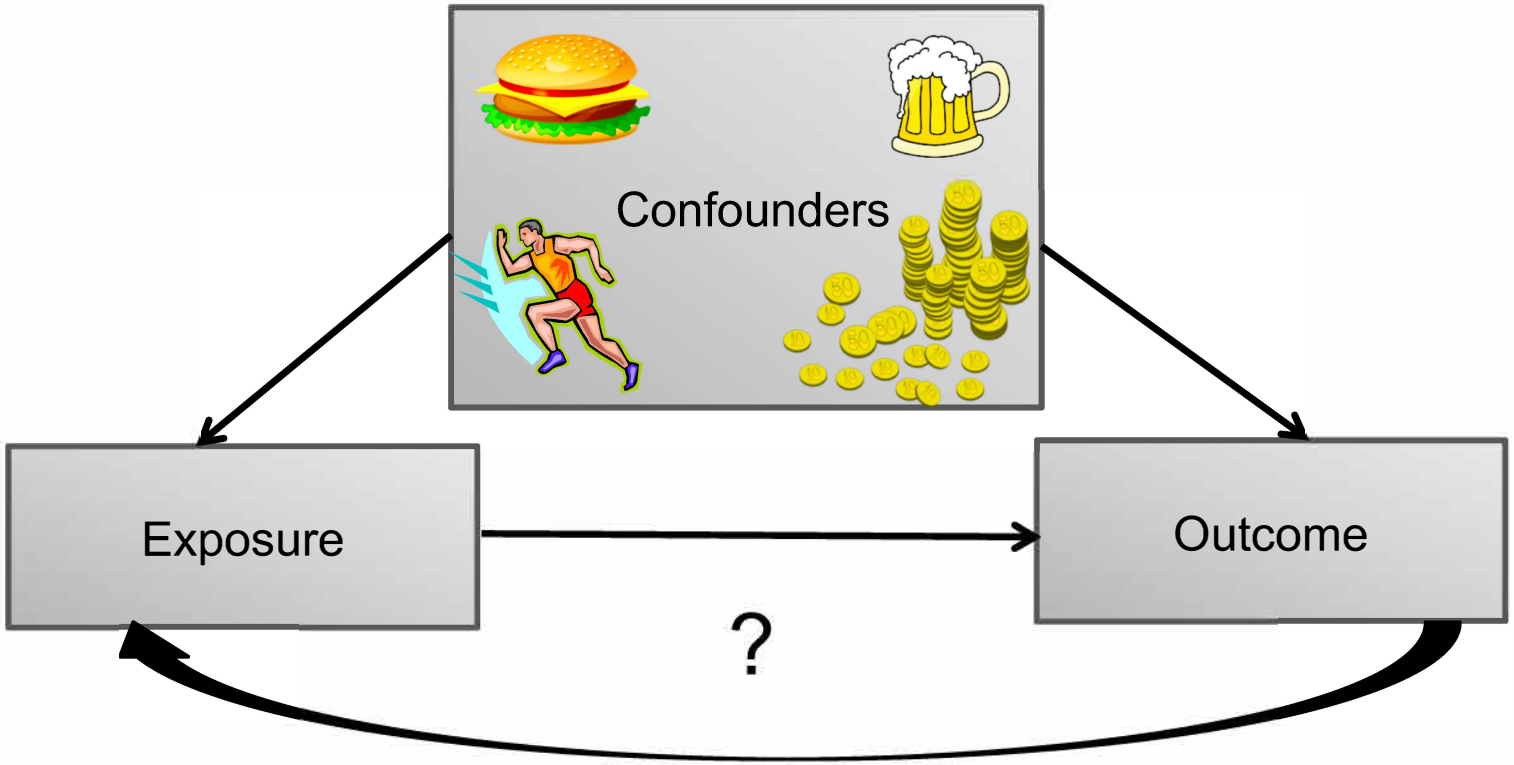
Causality

Animal studies

Qualitative studies

And more!

Conventional Epidemiology



Triangulation

Longitudinal studies

Negative controls

Causality

Cross-contextual

Mendelian randomization

Twin/family studies

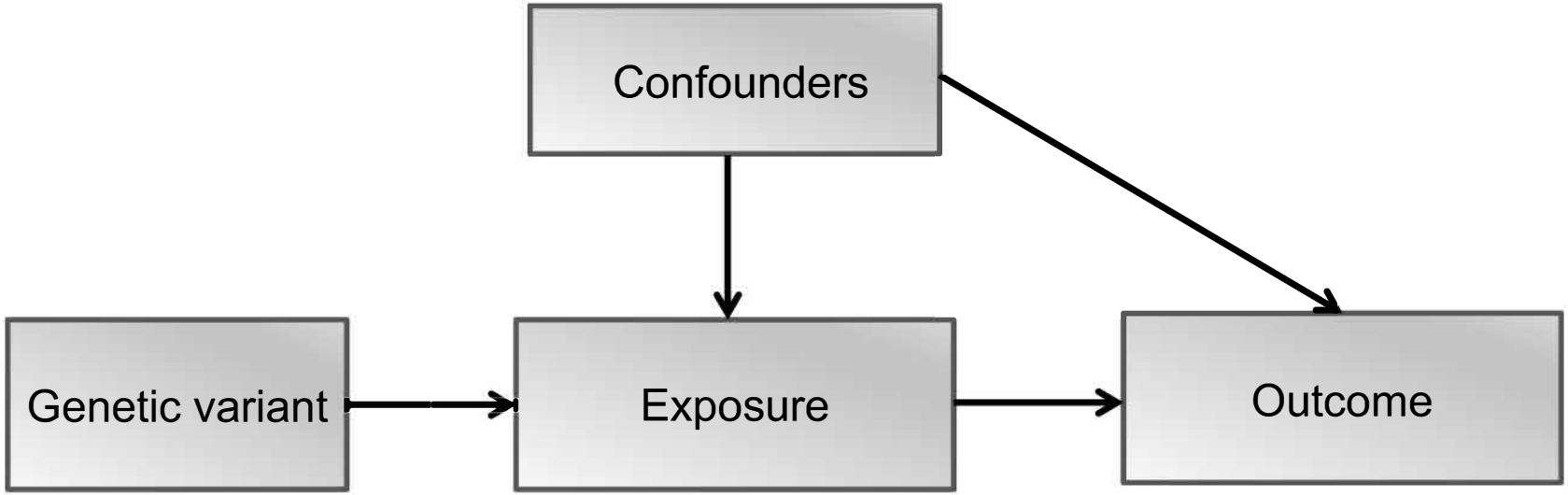
Gage et al. 2016 Annual Review of Psychology

Mendelian Randomization

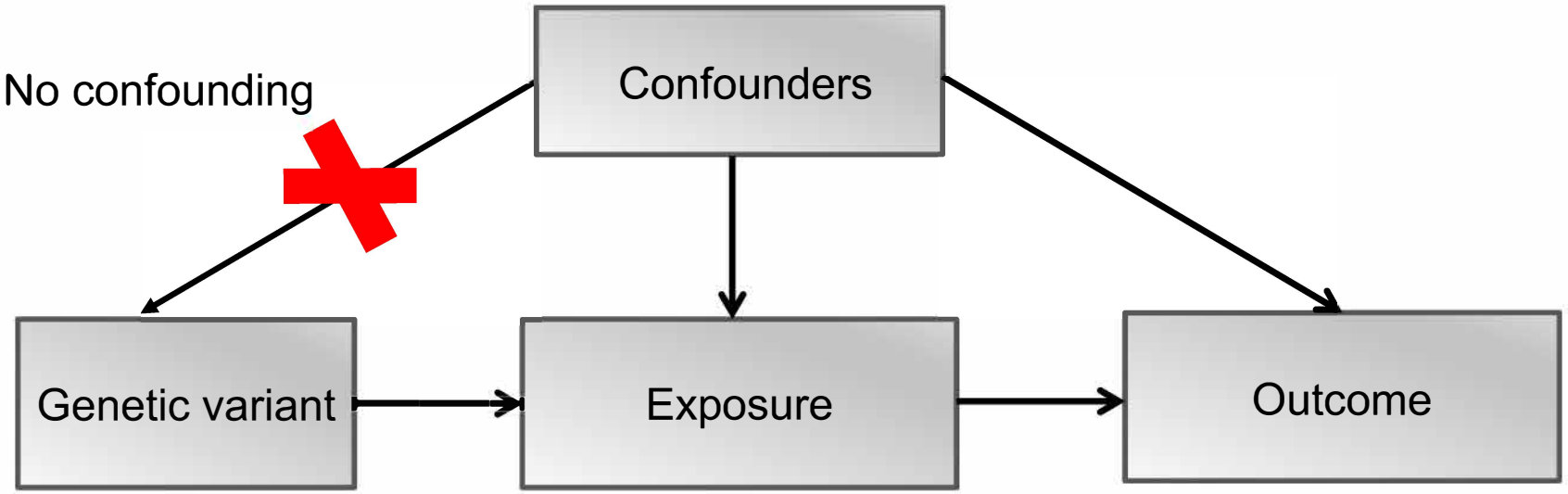
- Instrumental variable analysis
- Genetic variant is instrument
- Alleles inherited independently of confounding factors
- Genetic variants should not be affected by reverse causality

Davey Smith and Ebrahim, Int J Epidemiol. 2003

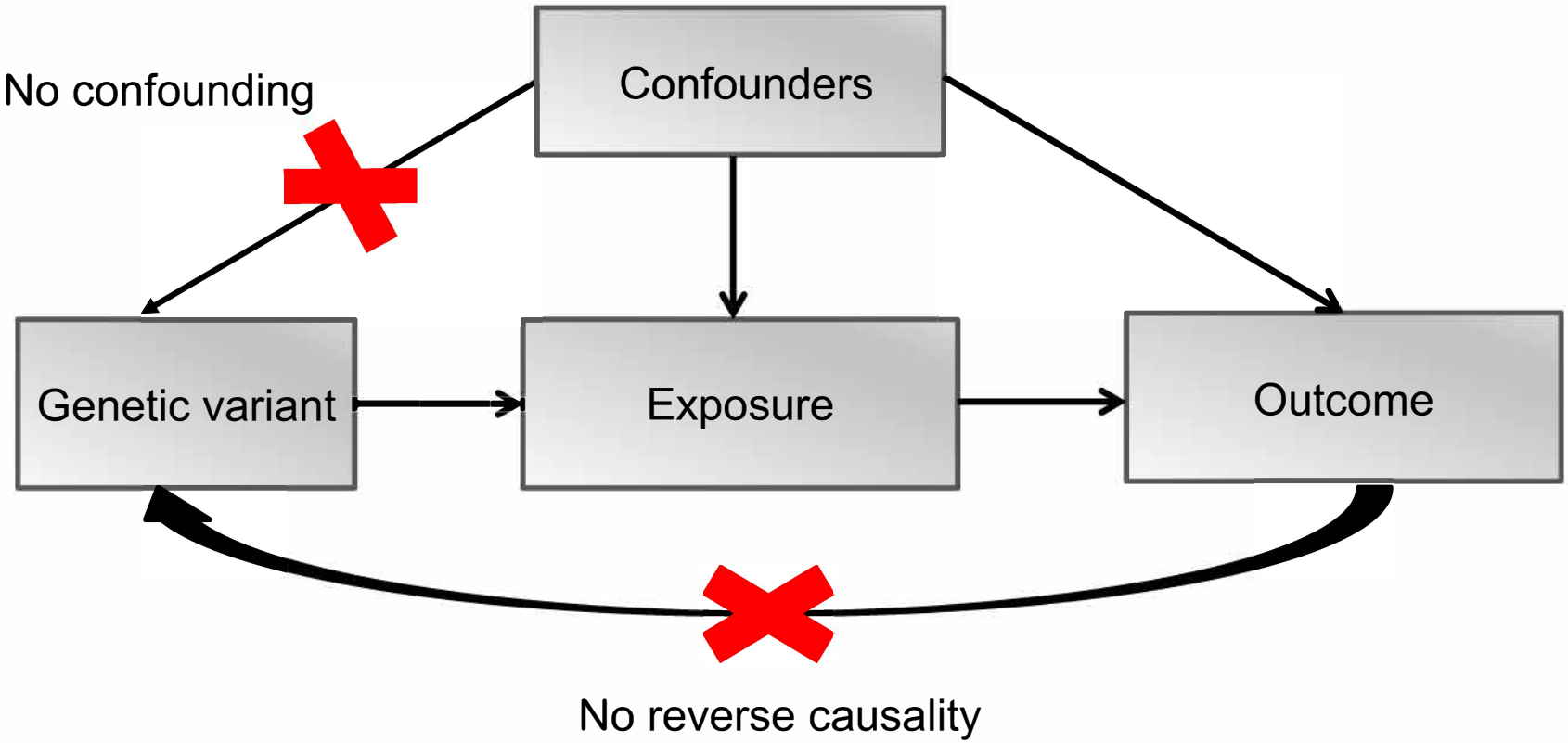
Mendelian Randomization

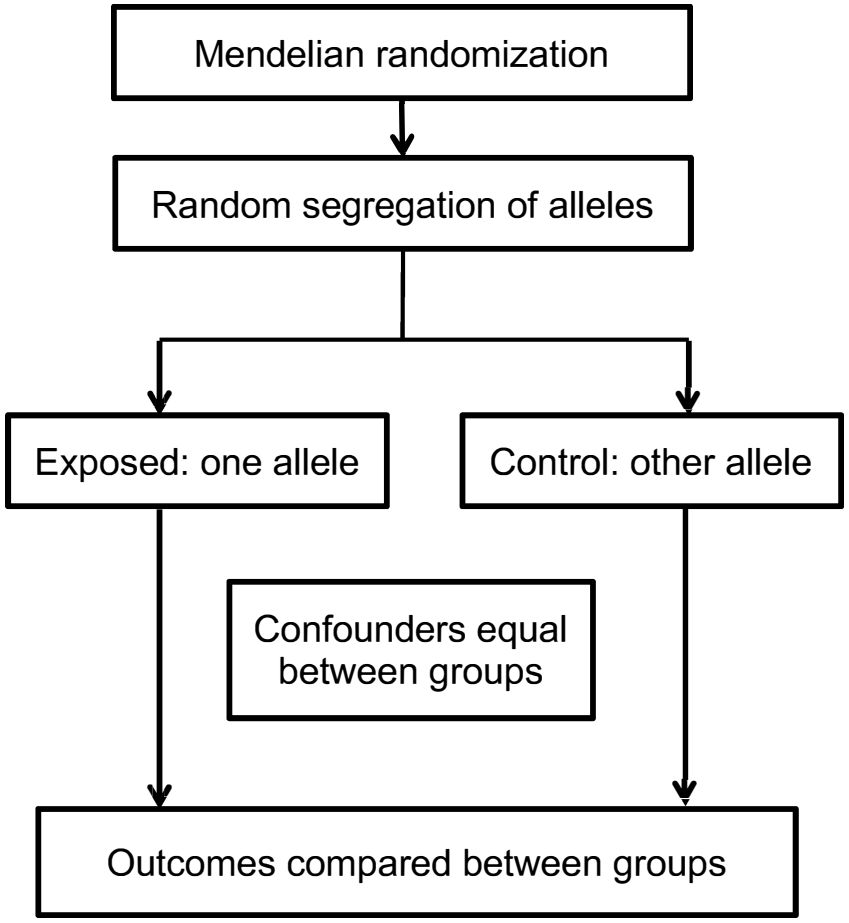


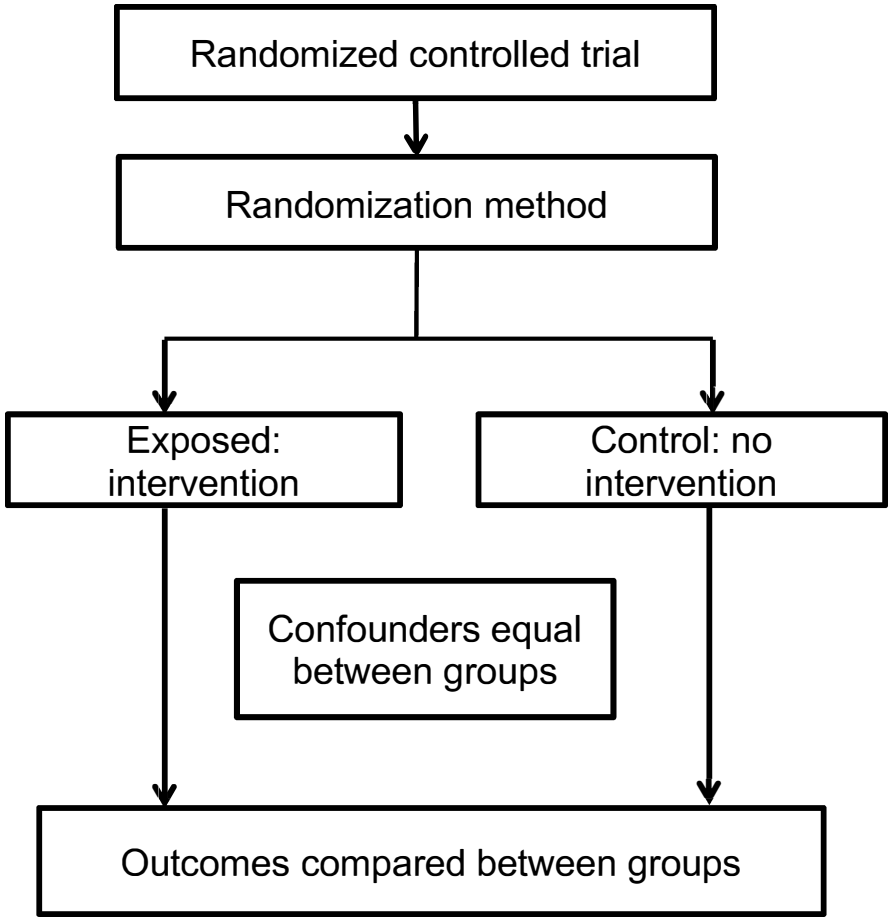
Mendelian Randomization



Mendelian Randomization







Davey Smith and Ebrahim, BMJ 2005;330:1076–9

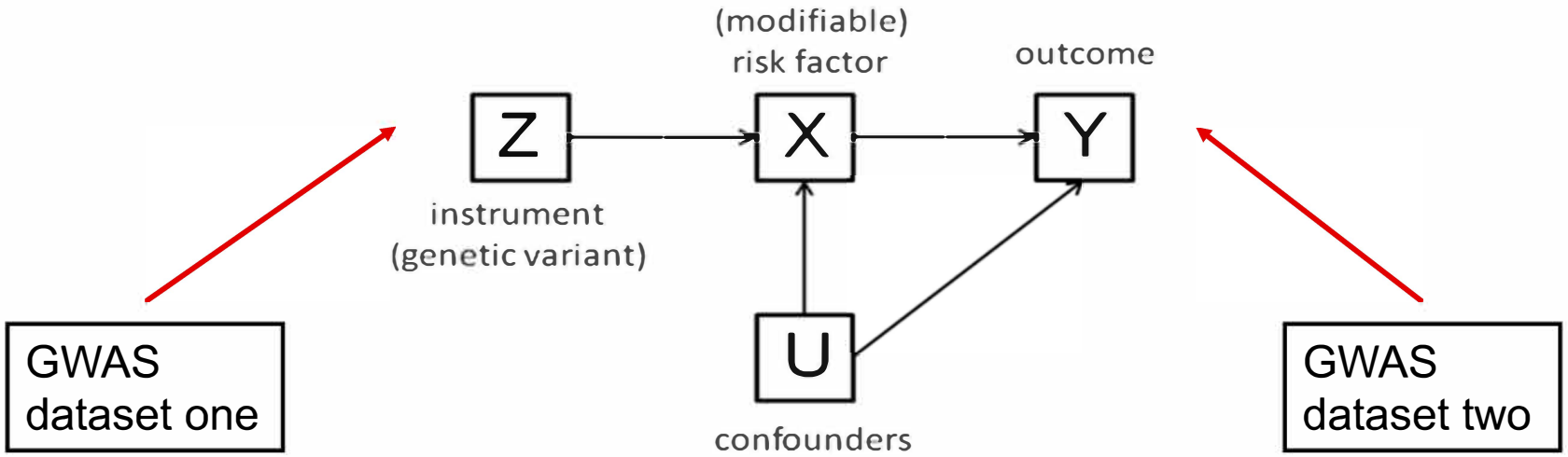
Assumptions of Mendelian Randomization

1) The genetic variant should be *reliably* associated with the exposure

2) The genetic variant should **only** be associated with the outcome through the exposure of interest

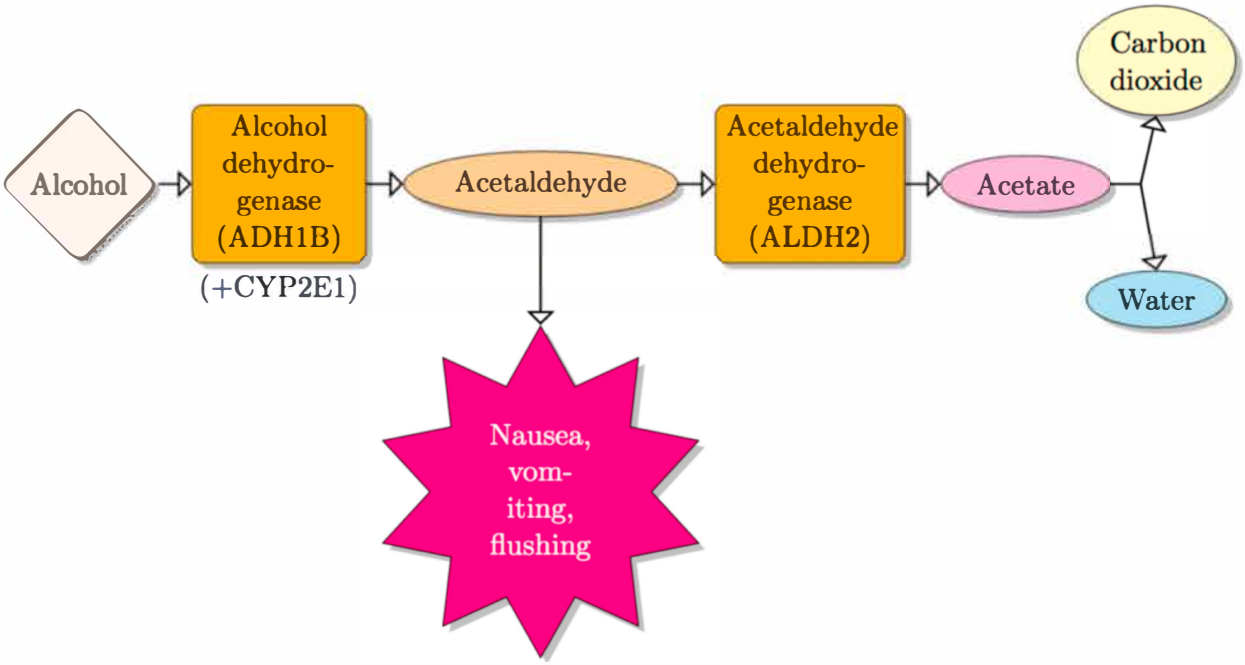
3) The genetic variant should be independent of other factors affecting the outcome (confounders)

Two-sample Mendelian randomization



GWAS: genomewide association study

Genetic variants for alcohol use



From: Commentary: Mendelian randomization-inspired causal inference in the absence of genetic data
Int J Epidemiol. 2016;46(3):962-965. doi:10.1093/ije/dyw327

Genetic variants predict alcohol use

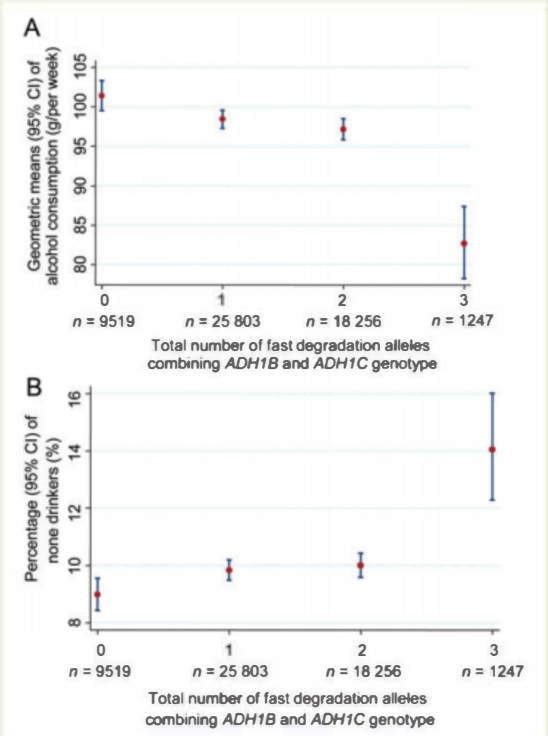
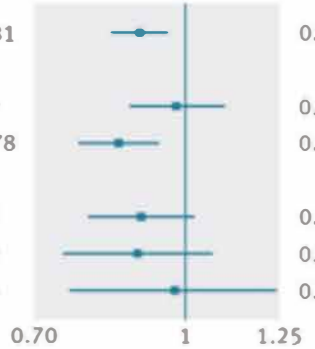


Figure 2 Association of combined *ADH1B* and *ADH1B* fast-allele score with average alcohol consumption levels in participants reporting drinking some alcohol and with not drinking alcohol; $n = 54\,604$. Shows geometric means (dots) and 95% confidence intervals of geometric means (vertical lines) of alcohol grams per week in those drinking some alcohol (*A*) and percentages (dots) and 95% confidence intervals of non-drinkers (*B*) by total number of fast-alleles.

From Lawlor et al (2013) Eur Heart J, 34(32); 2519-28

Genetic variants predict CHD

Category for coronary heart disease outcome	No of studies	No of cases/ individuals	Odds ratio (95% CI)	P value
Overall (all individuals)	46	20 259/168 731	0.90 (0.84 to 0.96)	0.001
Any or no alcohol intake				
Non-drinkers	31	5883/43 029	0.98 (0.88 to 1.10)	0.095*
Drinkers only	40	10 130/107 478	0.86 (0.78 to 0.94)	
Drinkers subgroup (units/week)				
Light (>0 to <7)	32	4686/47 246	0.90 (0.79 to 1.02)	0.828*
Moderate (≥7 to <21)	32	3222/33 772	0.89 (0.75 to 1.06)	
Heavy (≥21)	29	1919/16 225	0.97 (0.76 to 1.24)	



* P value for heterogeneity obtained from test for trend using meta-regression

Fig 2 Meta-analysis pooled estimates of the association between *ADH1B* rs1229984 (A-allele carriers *v* non-carriers) and coronary heart disease overall, and stratified by alcohol intake

From Holmes et al (2014) BMJ, 349;g4164

Mendelian randomization limitations

- Needs large sample size
- Assumption violations (hard to test)
- Needs relevant genetic variants
- Biological pleiotropy (one variant having multiple effects)
- Population stratification

Negative/positive controls

…compare analysis of interest with one where exposure (or outcome, not both!) changed to one with implausible (negative) or known (positive) effect, but where confounding likely to be the same

Negative/positive controls

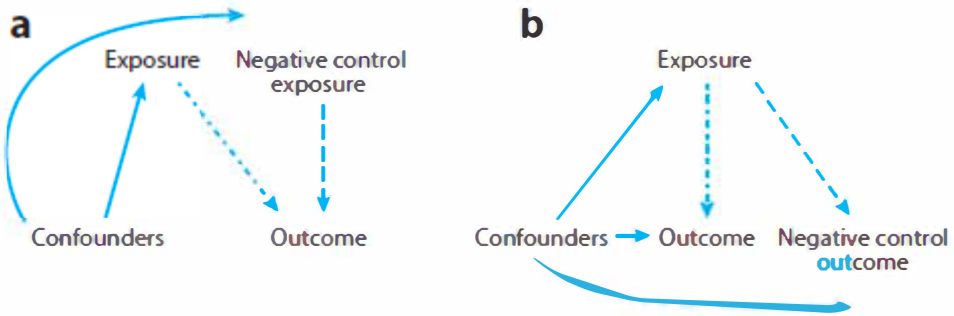


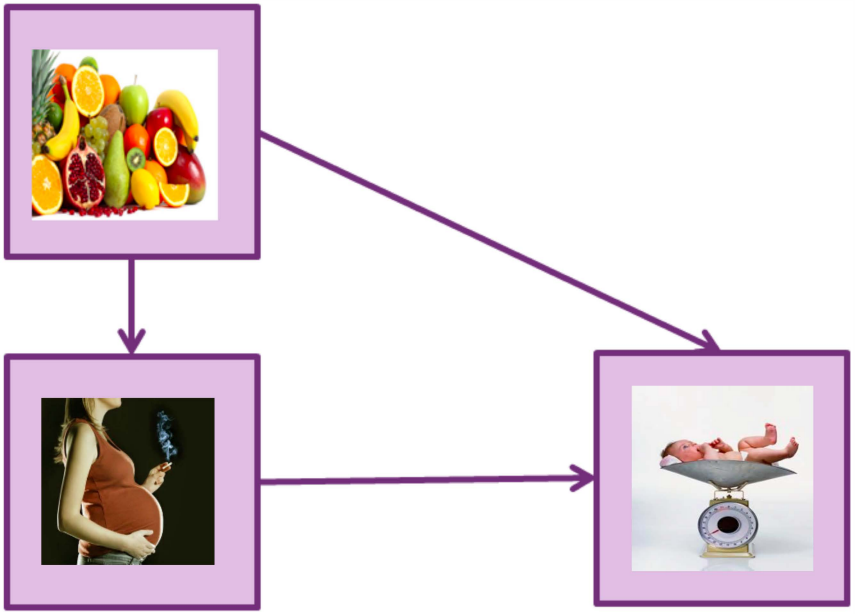
Figure 1

Schematic representations of (*a*) negative control exposure and (*b*) negative control outcome. Confounding is the same for the exposure or outcome and its negative control. However, there is no causal association between (*a*) the negative control exposure and the outcome of interest or (*b*) the exposure of interest and the negative control outcome. The dashed line represents the negative control analysis, and the dotted-and-dashed line represents the association under interrogation.

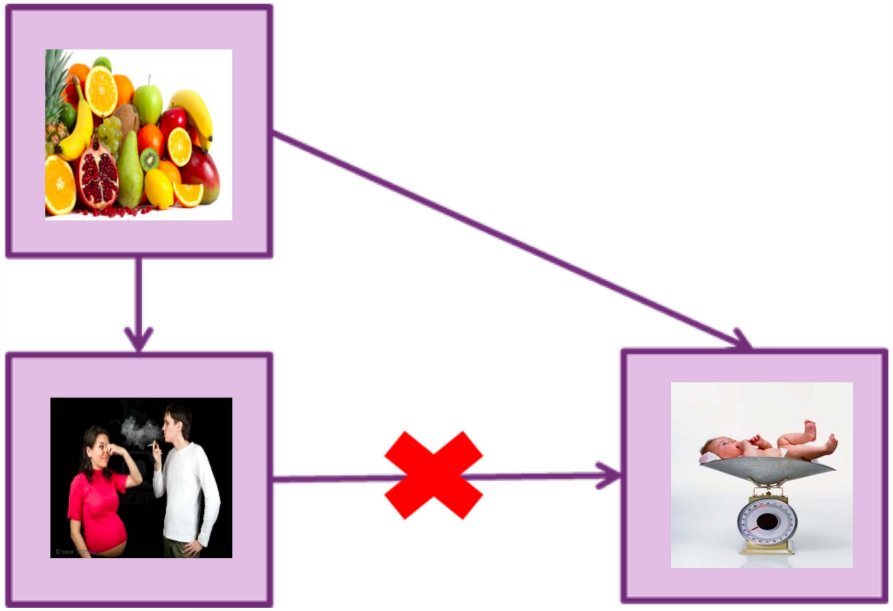
Gage et al, Annual Review of Psychology, 2016

Negative controls

1) Negative control



Intrauterine
tobacco exposure



Little/no
intrauterine
tobacco exposure

Negative/positive control examples

- Exposures
  - Comparing folate supplements during pregnancy and autism with fish oil supplements during pregnancy and autism (Suren et al, 2013)
    - Folate and fish oil supplements similarly confounded
    - Association between folate and autism, NOT fish oil and autism
    - Suggests causality rather than confounding
  - Comparing maternal smoking during pregnancy and offspring blood pressure with paternal smoking during partner pregnancy and BP (Brion et al, 2007).
    - Associations similar for maternal/paternal smoking.
    - Residual confounding?

Negative/positive control examples

- Outcomes
  - HRT and mortality from cardiovascular disease, compared to HRT and mortality from accidents, suicide and homicide (Pettiti et al, 1986/1987)
    - HRT use predicts lower cardiovascular mortality
    - BUT also predicts lower mortality from other reasons with no plausible biological mechanism
    - HRT use associated with lifestyle, socioeconomic, behavioural factors
    - Borne out in RCT – HRT cardiovascular disease association SPURIOUS

Negative/positive control limitations

- Associations could still be confounded by other factors not shared with the negative control
- Careful selection of negative/positive control required
  - Could be plausible causality with negative control (eg paternal smoking during pregnancy)
  - Positive control association could be due to confounding (cannabis and education?)

Cross contextual comparison

Association more likely to be causal if it is seen across different populations with different underlying confounding

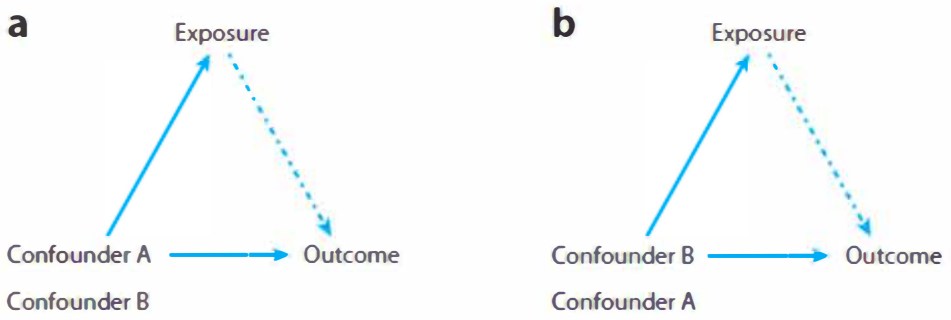


Figure 2

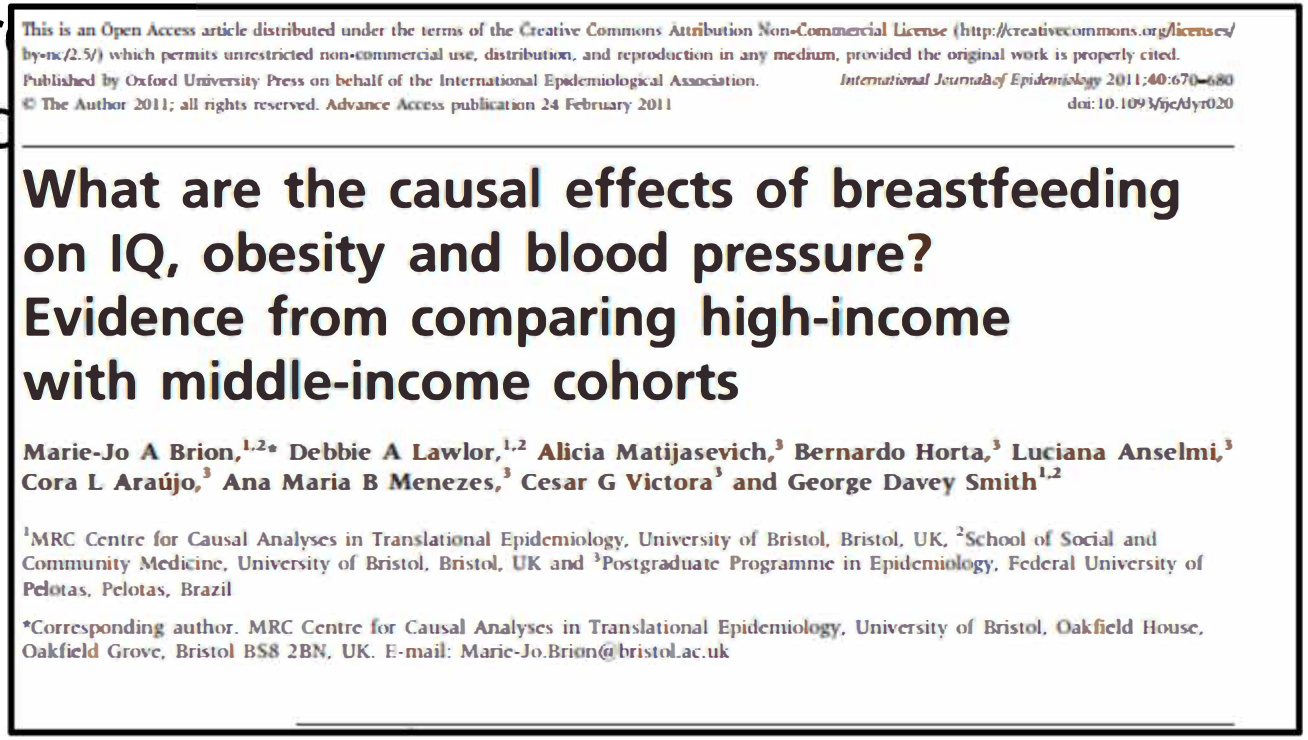
Schematic representations of a cross-contextual design. The exposure and outcome should be equivalent across the different contexts, but the confounding structure should not. Here, confounder A affects the relationship in context (*a*) but not in context (*b*). The reverse is true for confounder B.

Cross contextual comparison

Association more likely to be causal if it is seen across different populations with different underlying confounding

Cross contextual comparison

Association more likely to be causal if it is seen acr
und



Published by Oxford University Press on behalf of the International Epidemiological Association. *International Journal of Epidemiology* 2011;40:670–680
© The Author 2011; all rights reserved. Advance Access publication 24 February 2011 doi:10.1093/ije/dyr020

What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts

Marie-Jo A Brion,[1,2]* Debbie A Lawlor,[1,2] Alicia Matijasevich,[3] Bernardo Horta,[3] Luciana Anselmi,[3] Cora L Araújo,[3] Ana Maria B Menezes,[3] Cesar G Victora[3] and George Davey Smith[1,2]

[1]MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK, [2]School of Social and Community Medicine, University of Bristol, Bristol, UK and [3]Postgraduate Programme in Epidemiology, Federal University of Pelotas, Pelotas, Brazil

*Corresponding author. MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. E-mail: Marie-Jo.Brion@bristol.ac.uk

Brion et al., Int J Epidemiol 2008

Table 1 Distribution of infants according to duration of any breastfeeding

Breastfeeding duration (months)	Prevalence (%)	
	Pelotas	ALSPAC
0 to <1	15.6	36.8
1 to <3	25.4	15.6
3 to <6	23.6	13.7
⩾6	35.3	33.9

For analyses, the categories of never breastfed and breastfed <1 month were merged as the prevalence of never breastfed in Pelotas was extremely low and there is substantial misclassification between these categories.[11]

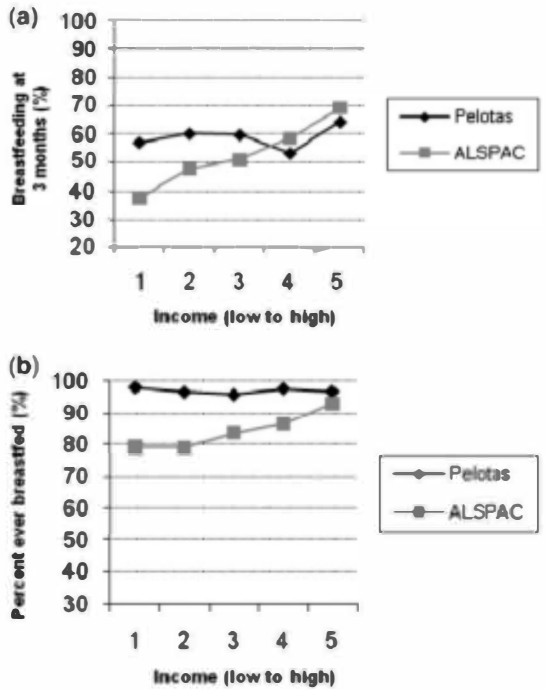
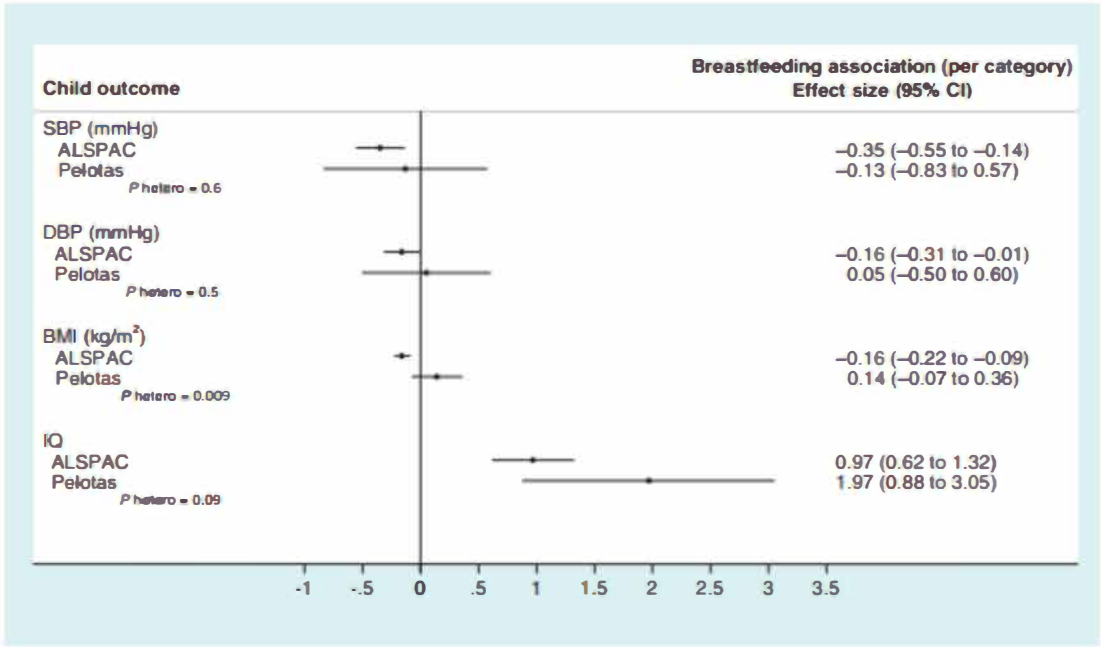


Figure 1 (a) Prevalence of breastfeeding (exclusive or non-exclusive) at 3 months by family income group and (b) prevalence of ever breastfed by family income



Brion et al., Int J Epidemiol 2008

Cross contextual comparison

Table 4 Summary of results from the cross-cohort comparison and validation using a randomized trial

	Comparison method						Validation	
	Association with any breastfeeding (per category)[a]						Effect of breastfeeding intervention	
	ALSPAC			Pelotas			Belarus	
	Strong socio-economic patterning in breastfeeding			Weak socio-economic patterning in breastfeeding			Randomized trial	
Outcome	β	95% CI	*P*	β	95% CI	*P*	Difference in outcome[b]	95% CI
SBP (mmHg)	−0.35	−0.55 to −0.14	0.001	−0.03	−0.83 to 0.57	0.7	0.2	−2.9 to 3.3
DBP (mmHg)	−0.16	−0.31 to −0.01	0.04	0.05	−0.50 to 0.60	0.9	0.2	−1.8 to 2.2
BMI (kg/m^2)	−0.16	−0.22 to −0.09	<0.001	0.14	−0.07 to 0.36	0.2	0.1	−0.2 to 0.3
IQ	0.97	0.62 to 1.32	<0.001	1.97	0.88 to 3.05	<0.001	5.9	−1.0 to 12.8

[a]None/<1 month; 1 to <3 months; 3 to <6 months; ⩾6 months; fully adjusted models.
[b]Intervention vs control. Results extracted from publications from the Belarus PROBIT trial.[16,17]
CI, confidence interval.

Brion et al., Int J Epidemiol 2008

Cross contextual study limitations

- Is underlying confounding definitely different?
  - Similar confounders between the two will render design inappropriate
- Harmonization of exposure and outcome between contexts
- Still possibility that different confounding in each context is still influencing both results

Sibling and twin studies

- Discordant twin studies
  - One has outcome of interest, other does not
  - Ideally matched pairs for case-control design
- Monozygotic versus dizygotic twins
- Sibling/cousin pairs

Sibling and twin studies limitations

- Hard to find discordant twins
- Even twins have some different environmental impacts!
- Lack of generalisability – intrauterine experience
- Are identical twins treated more similarly than non-identical twins?

Other designs: natural experiments

- Dutch Hunger Winter
- China – Great Leap Forward
- Local policy changes (eg age of compulsory education changes, alcohol policy changes)

- Limitations
  - Can’t predict/plan
  - Can’t control for other factors
  - Extreme situations (famine/war) could have other impacts

Summary

- Mendelian randomization
  - Unconfounded(?) genetic proxy for exposure
- Negative/positive controls
  - Similar confounding between assoc. of interest and +ve/-ve control
- Cross-contextual studies
  - Assess association of interest in 2 datasets where confounding differs
- Twin and family studies
  - Discordant identical twins as matched cases/controls
- Natural experiments

Triangulation

Longitudinal studies

Negative controls

Causality

Cross-contextual

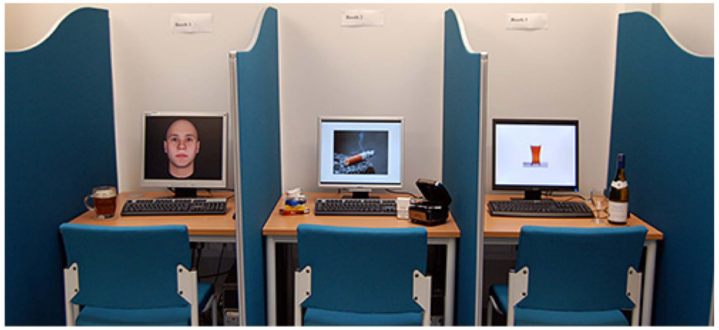
Mendelian randomization

Twin/family studies

Gage et al. 2016 Annual Review of Psychology

Triangulation

Observational studies



Experimental studies

Causality

Animal studies

Qualitative studies

And more!

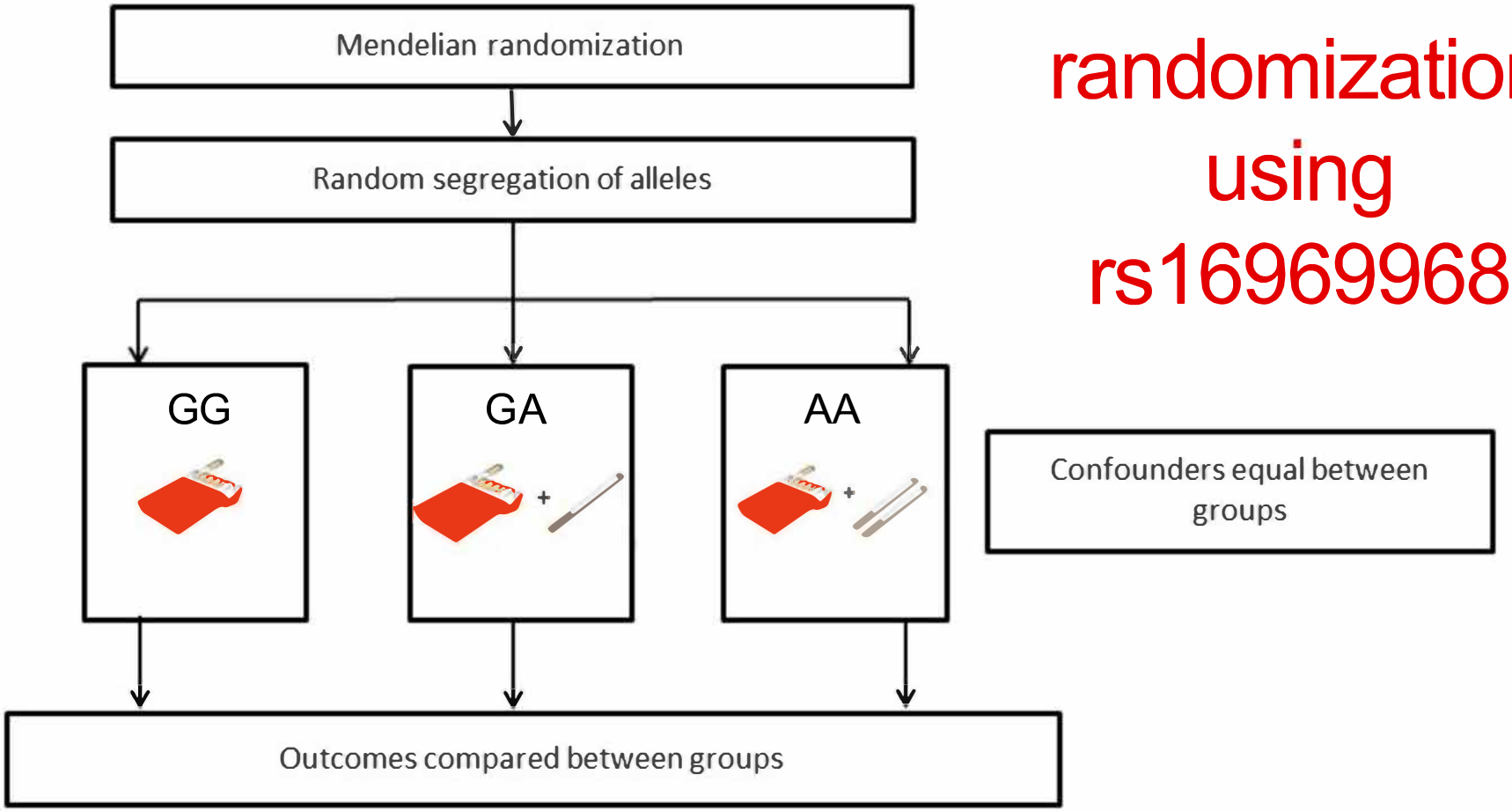
Acknowledgements

Thanks to:

George Davey Smith	University of Bristol, UK
Marcus Munafò	University of Bristol, UK
Amy Taylor	University of Bristol, UK
Stephen Burgess	University of Cambridge, UK

Contact: s.gage@liverpool.ac.uk; @soozaphone

Mendelian randomization using rs16969968



Examples: Smoking related diseases

- Lung cancer

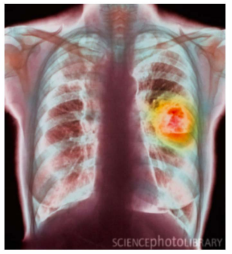
  (Amos, 2010; Lips, 2010; Spitz, 2008)

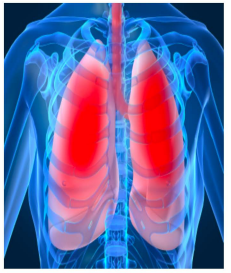
- COPD/emphysema

  (Kaur-Knudsen, 2011; Lambrechts, 2010; Pillai, 2009)

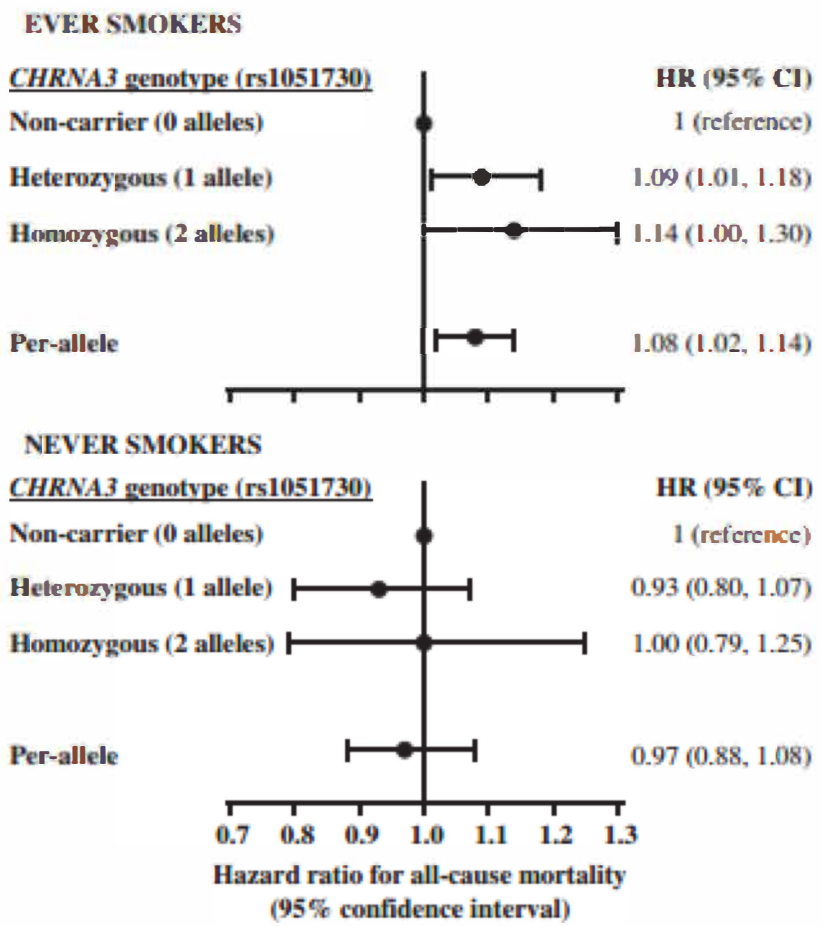
- Peripheral Arterial Disease

  (Thorgeirsson, 2008)





Examples: Smoking and Mortality



Rode et al., Int J Epidemiol 2014