SER 2020

A COMPARISON OF ANALYTICAL APPROACHES TO OBTAIN MENDELIAN RANDOMIZATION ESTIMATES WITH LONGITUDINAL EXPOSURES

JOY SHI, PHD
POSTDOCTORAL FELLOW
DEPARTMENT OF EPIDEMIOLOGY
HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

MOTIVATIONS.

MENDELIAN RANDOMIZATION

- Genetic variants are proposed as an instrumental variable (IV)
- Increasingly popular approach to causal inference

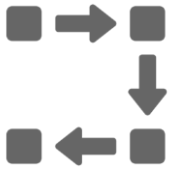
CURRENT LIMITATIONS

- Exposures of interest are often time-varying,
  e.g. alcohol, cholesterol levels
- Conventional IV methods designed to only handle time-fixed exposures
- MR estimates based on a single measurement of the exposure

OBJECTIVES.

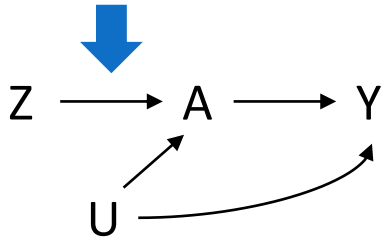
Describe the instrumental conditions in the context of a time-varying exposure

Review the possible causal interpretations of MR estimates and the additional assumptions required to do

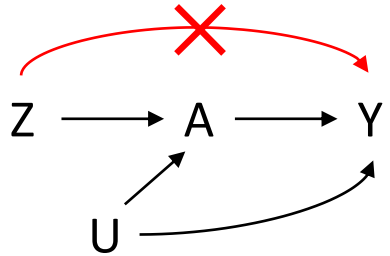


Propose methods to incorporate time-varying exposures in MR analyses of continuous outcomes based on g-estimation of structural mean models

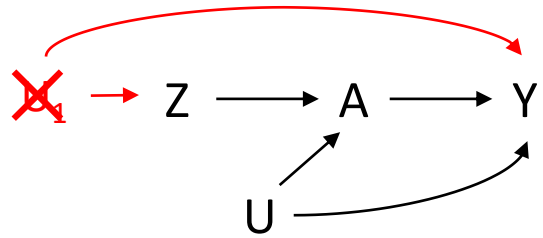
IV CONDITIONS.



(1) Genetic variant is associated with the exposure.



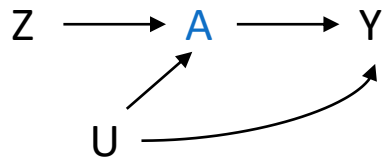
(2) The genetic variant does not affect the outcome except through its potential effect on the exposure.



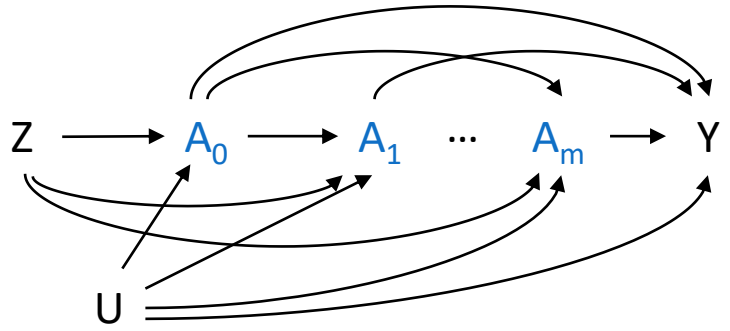
(3) The genetic variant and the outcome do not share common causes.

Z = Genetic variant
A = Exposure
Y = Outcome
U = (Unmeasured) confounders

IV CONDITIONS WITH A TIME-VARYING EXPOSURE.



TIME-FIXED EXPOSURE



TIME-VARYING EXPOSURE

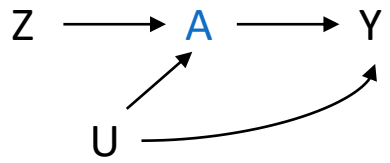
Z = Genetic variant

A = Exposure

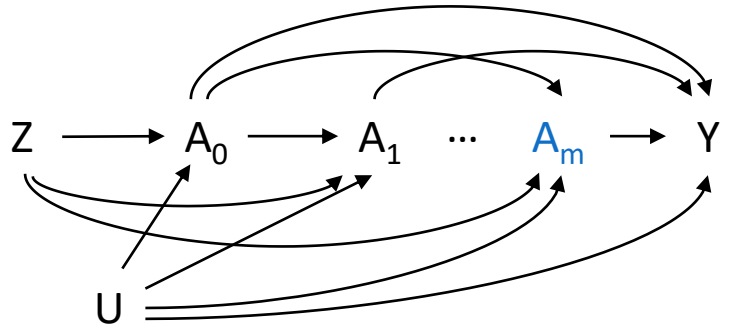
Y = Outcome

U = (Unmeasured) confounders

IV CONDITIONS WITH A TIME-VARYING EXPOSURE.



TIME-FIXED EXPOSURE

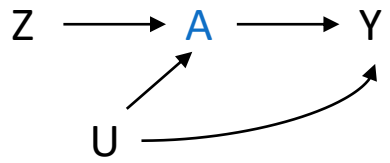


TIME-VARYING EXPOSURE

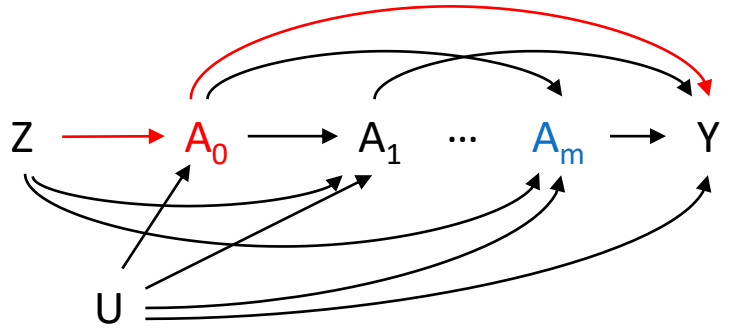
Z = Genetic variant
A = Exposure
Y = Outcome
U = (Unmeasured) confounders

Second IV condition: the genetic variant does not affect the outcome except through its possible effect on the exposure

IV CONDITIONS WITH A TIME-VARYING EXPOSURE.



TIME-FIXED EXPOSURE



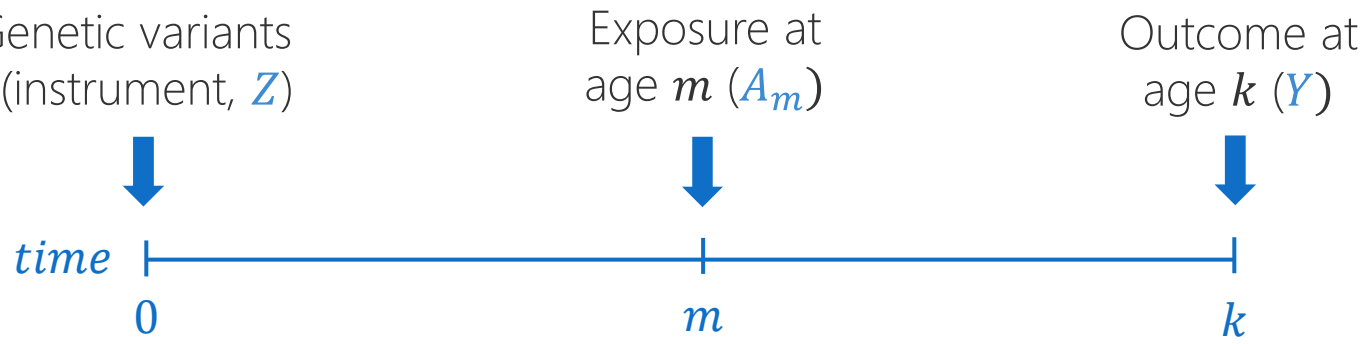
TIME-VARYING EXPOSURE

Z = Genetic variant
A = Exposure
Y = Outcome
U = (Unmeasured) confounders

Recall: the second IV condition states that the instrument does not affect the outcome except through its possible effect on the exposure

FIRST TYPE OF CAUSAL EFFECT: THE POINT EFFECT.

Suppose for each individual, we have measurements on:



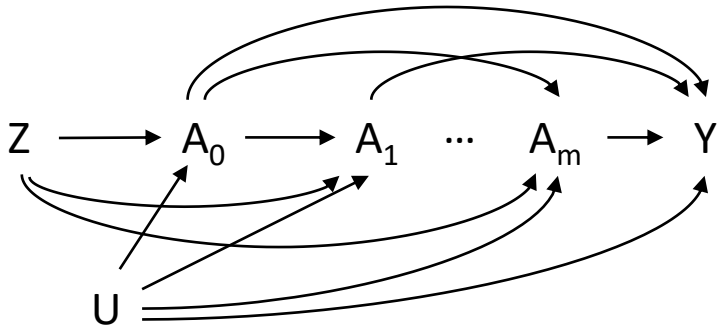
Point effect:

$$\mathrm{E}[Y^{a_m}] - \mathrm{E}[Y^{a'_m}]$$

Outcome that an individual would have had had they received
exposure level a at age m (Y^{a_m})
exposure level a' at age m ($Y^{a'_m}$)

FIRST TYPE OF CAUSAL EFFECT: THE POINT EFFECT.

Requires genetic variant Z to be a valid instrument for A_m (and only A_m)

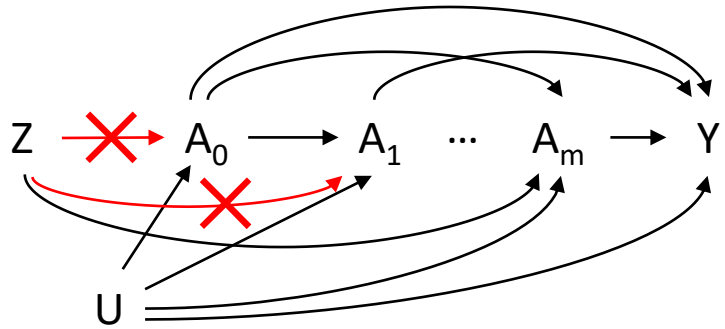


Point effect: $\mathrm{E}[Y^{a_m}] - \mathrm{E}[Y^{a'_m}]$

FIRST TYPE OF CAUSAL EFFECT: THE POINT EFFECT.

Requires genetic variant Z to be a valid instrument for A_m (and only A_m)

- No arrows from Z to other exposure time points; *or*

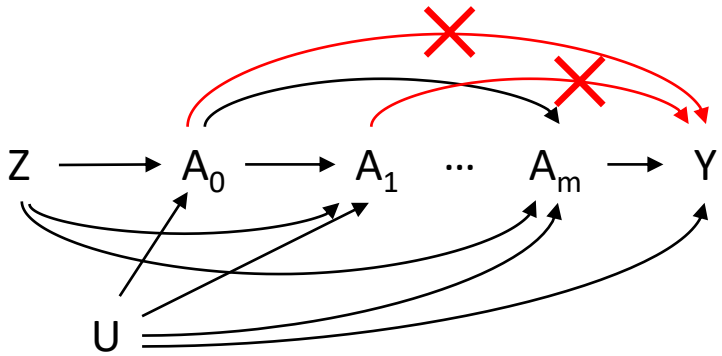


Point effect: $\mathrm{E}[Y^{a_m}] - \mathrm{E}[Y^{a'_m}]$

FIRST TYPE OF CAUSAL EFFECT: THE POINT EFFECT.

Requires genetic variant Z to be a valid instrument for A_m (and only A_m)

- No arrows from Z to other exposure time points; *or*
- No arrows from other exposure time points to the outcome

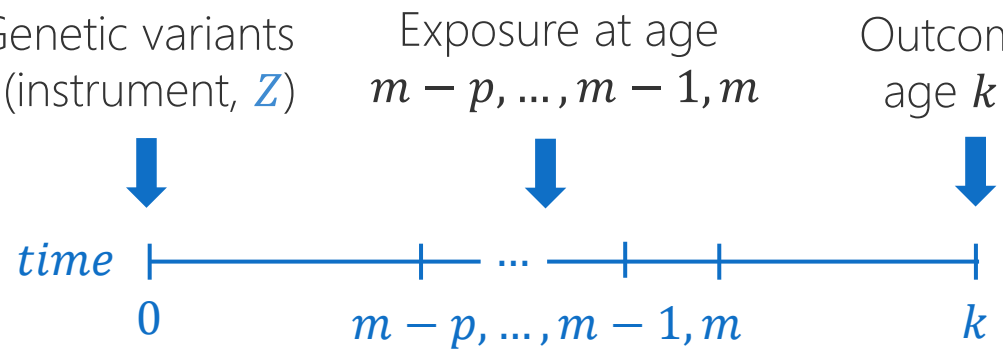


The assumptions to estimate the point effect in MR studies generally do not hold.

Point effect: $\mathrm{E}[Y^{a_m}] - \mathrm{E}[Y^{a'_m}]$

SECOND TYPE OF CAUSAL EFFECT: THE PERIOD EFFECT.

Suppose we are now interested in the effect of an exposure during a specific period, e.g. $[m-p, \ldots, m-1, m]$



Interested counterfactual outcomes under:

1. Some exposure trajectory $(a_{m-p}, \ldots, a_{m-1}, a_m)$
2. Same exposure trajectory as (1), but shifted upwards by 1 unit

Period effect:

$$\mathrm{E}[Y^{a_{m-p}+1,\ldots,a_{m-1}+1,\, a_m+1}] - \mathrm{E}[Y^{a_{m-p},\ldots,a_{m-1},a_m}]$$

SECOND TYPE OF CAUSAL EFFECT: THE PERIOD EFFECT.

If we measure the period effect with a single measurement time point:

1. The genetic variant Z is a valid instrument for the exposure during the period $[m-p, \ldots, m-1, m]$
2. The association between the genetic variant and the exposure stays constant over time

This is a weaker form of the assumption required for the point effect

With multiple measurements of an exposure, we can relax this assumption by allowing the association to change over time.

For the period effect:

$$\mathrm{E}\left[Y^{a_{m-p}=a,\ldots,a_{m-1}=a,\,a_m=a} - Y^{\bar{0}} \middle| A_{m-p},\ldots,A_{m-1},A_m,Z\right] = \gamma(a;\psi)$$

STRUCTURAL MEAN MODELS.

Model for the expected difference between two counterfactual outcomes

STRUCTURAL MEAN MODELS.

For the period effect:

$$\mathrm{E}\left[Y^{a_{m-p}=a,\ldots,a_{m-1}=a,\,a_m=a} - Y^{\bar{0}} \mid A_{m-p},\ldots,A_{m-1},A_m,Z\right] = \gamma(a;\psi)$$

Model for the expected difference between two counterfactual outcomes

Function of the exposure, indexed by the unknown parameter ψ

E.g., with one measurement of the exposure, $\gamma(a;\psi) = \psi a$

with $p+1$ measurements of the exposure: $\gamma(a;\psi) = \sum_{j=m-p}^{p} \psi_j a$

Homogeneity: Z does not modify the relationship between the exposure and the outcome

G-ESTIMATION.

Define an estimating equation

$$U(\psi; \mathrm{Z}) = H(\psi)(\mathrm{Z} - \mathrm{E}(\mathrm{Z}))$$

where $H(\psi) = Y - \gamma(a; \psi)$

Set the estimating equation to 0 and solve for ψ

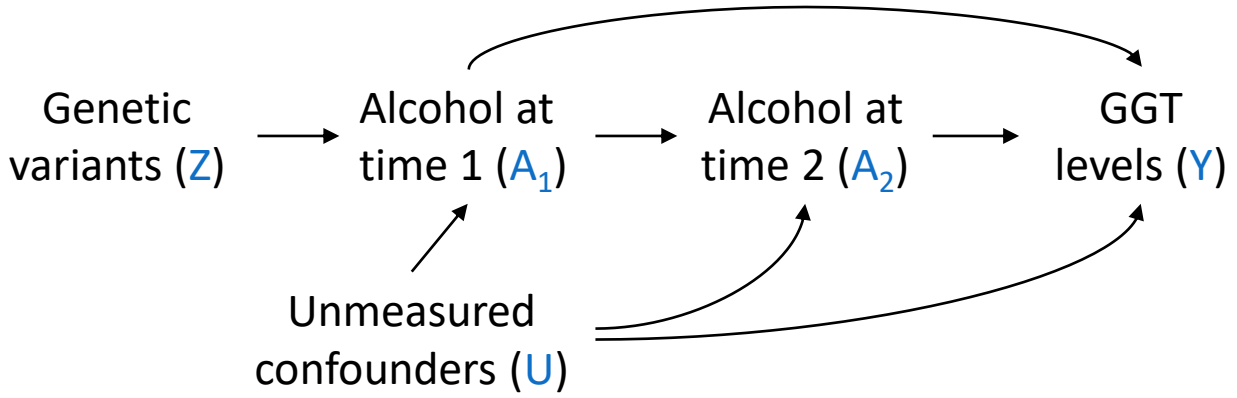
Set $E[U(\psi; Z)] = 0$ and solve for ψ; for continuous outcomes, the solution is:

$$\hat{\psi} = \sum_{i=1}^{n} Y_i(Z_i - \mathrm{E}[Z]) \,/\, \sum_{i=1}^{n} A_{m,i}(Z_i - \mathrm{E}[Z])$$

APPLICATION: OVERVIEW.

Goal: use data from the Framingham Heart Study (FHS) to estimate the period effect* of shifting one's alcohol intake trajectory upwards by 1 drink/day on gamma-glutamyl transferase (GGT) levels

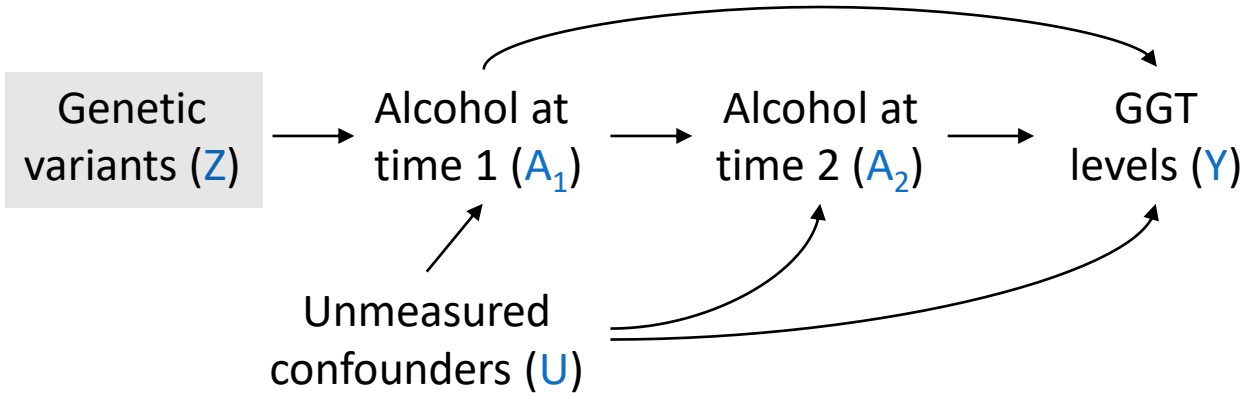
*The period was defined to be over the duration of follow-up



APPLICATION: OVERVIEW.

Goal: use data from the Framingham Heart Study (FHS) to estimate the period effect* of shifting one's alcohol intake trajectory upwards by 1 drink/day on gamma-glutamyl transferase (GGT) levels

*The period was defined to be over the duration of follow-up

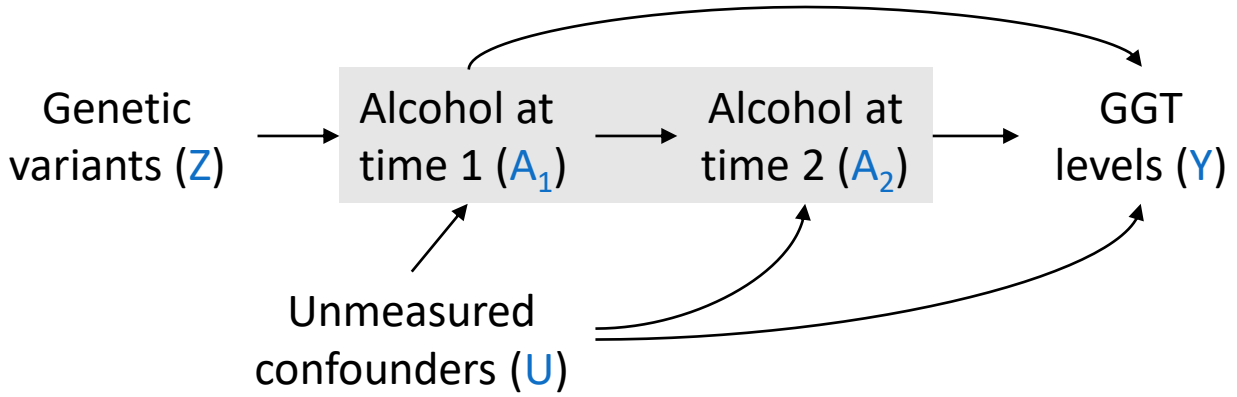


INSTRUMENT: 26 genetic variants identified from prior studies that were combined in different ways to create multiple weighted allele scores;

APPLICATION: OVERVIEW.

Goal: use data from the Framingham Heart Study (FHS) to estimate the period effect* of shifting one's alcohol intake trajectory upwards by 1 drink/day on gamma-glutamyl transferase (GGT) levels

*The period was defined to be over the duration of follow-up

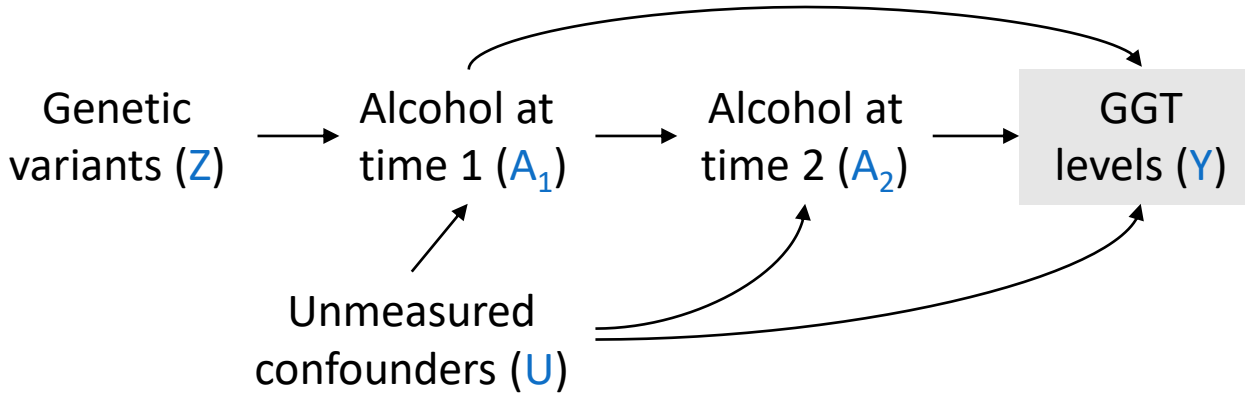


EXPOSURES: alcohol intake (via FFQ) at examination 1 and examination 2

APPLICATION: OVERVIEW.

Goal: use data from the Framingham Heart Study (FHS) to estimate the period effect* of shifting one's alcohol intake trajectory upwards by 1 drink/day on gamma-glutamyl transferase (GGT) levels

*The period was defined to be over the duration of follow-up



OUTCOME: gamma-glutamyl transferase levels (measured in blood) at examination 2; marker for liver function

APPLICATION: ANALYSIS.

Proposed two structural mean models

Only consider a single exposure measurement

1. Exposure measured at examination 1
2. Exposure measured at examination 2
3. Average exposure across the two examination time points

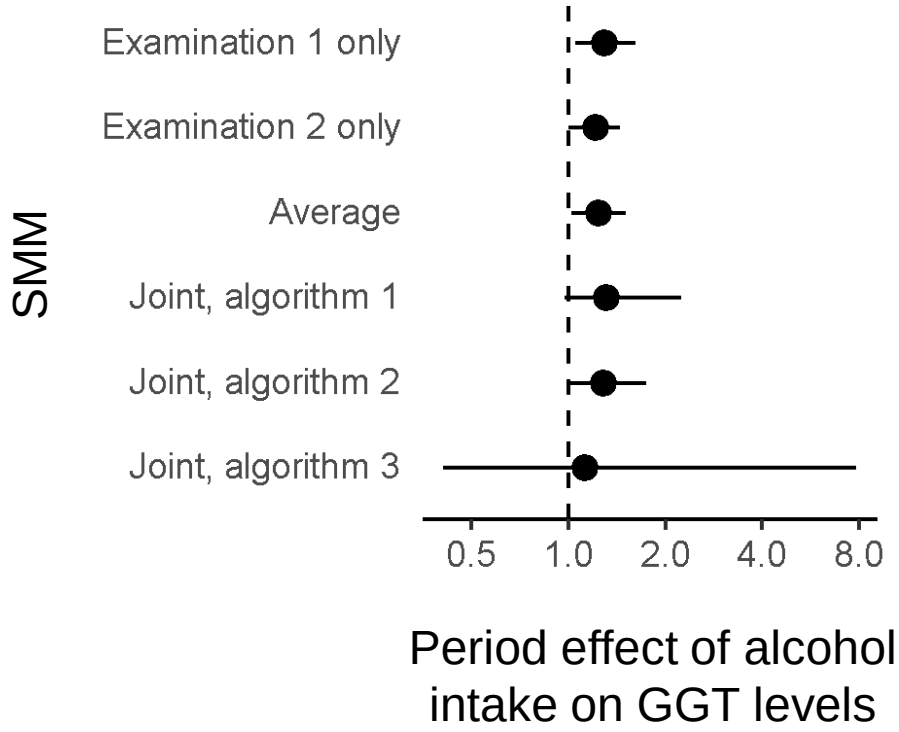
Consider multiple exposure measurements

Estimated the parameters using different weighted allele scores

$$\mathrm{E}\left[Y - Y^{\bar{0}} \middle| A_{m-p}, \ldots, A_{m-1}, A_m, Z\right] = \psi_1 A$$

$$\mathrm{E}\left[Y - Y^{\bar{0}} \middle| A_{m-p}, \ldots, A_{m-1}, A_m, Z\right] = \sum_{j=1}^{2} \psi_j A$$

APPLICATION: RESULTS.

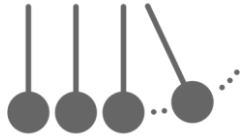


Highlights from our data application:

- Similarity of results driven by limited number of exposure measurements
- Precision of estimates

Takeaway: estimates were similar regardless of choice of model, although confidence intervals vary

SUMMARY.



Possible causal interpretations of MR estimates and their assumptions

- Point effect
- Period effect

G-estimation of structural mean models to accommodate multiple measures of a time-varying exposure in an MR analysis.

Advantages of our proposed method:

- Include >2 time points
- Model non-linear outcomes
- Include other covariates into the model

ACKNOWLEDGEMENTS.

Miguel Hernan

Sonja Swanson

Collaborators

- Peter Kraft
- Immaculata De Vivo
- Bernard Rosner

Funding

UCB Fellowship

HARVARD T.H. CHAN | SCHOOL OF PUBLIC HEALTH

QUESTIONS.