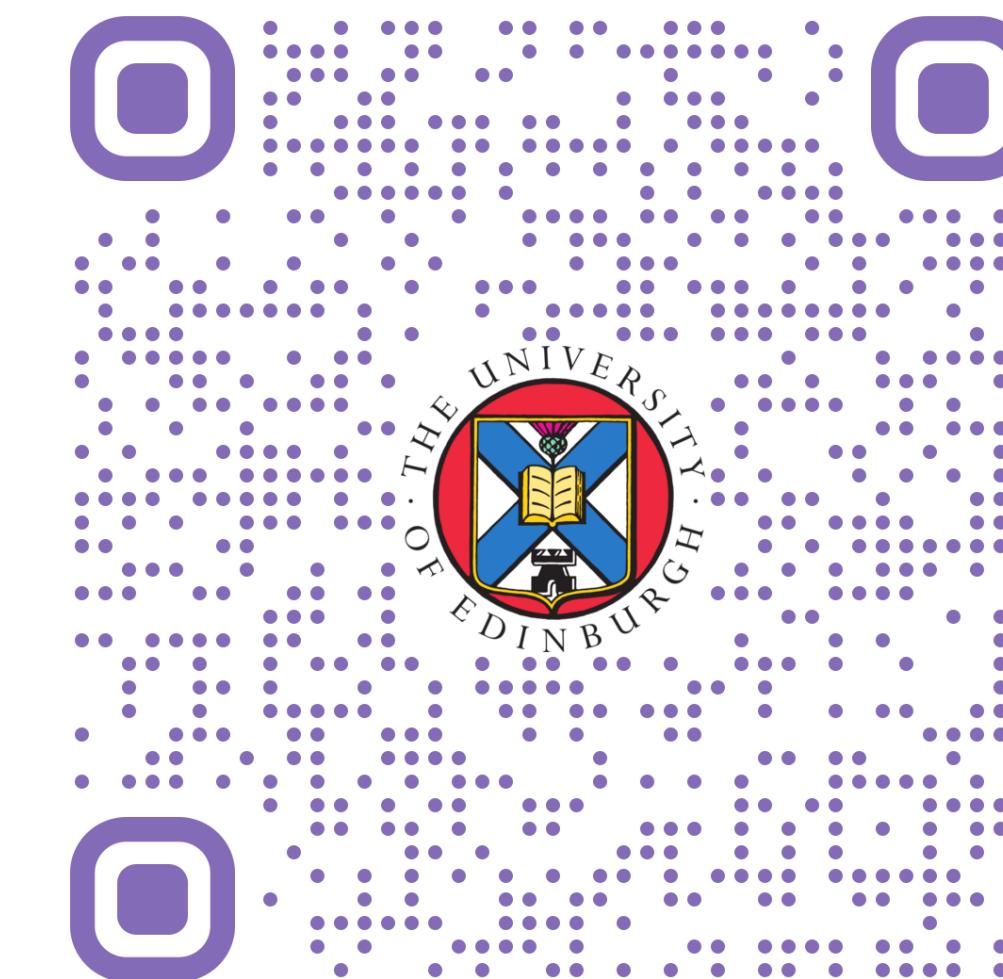


Christopher Aldous Oldnall | Thursday 6th July 2023
34th European Meeting of Statisticians, Contributed Session: Causal Inference 2

Mendelian randomisation:

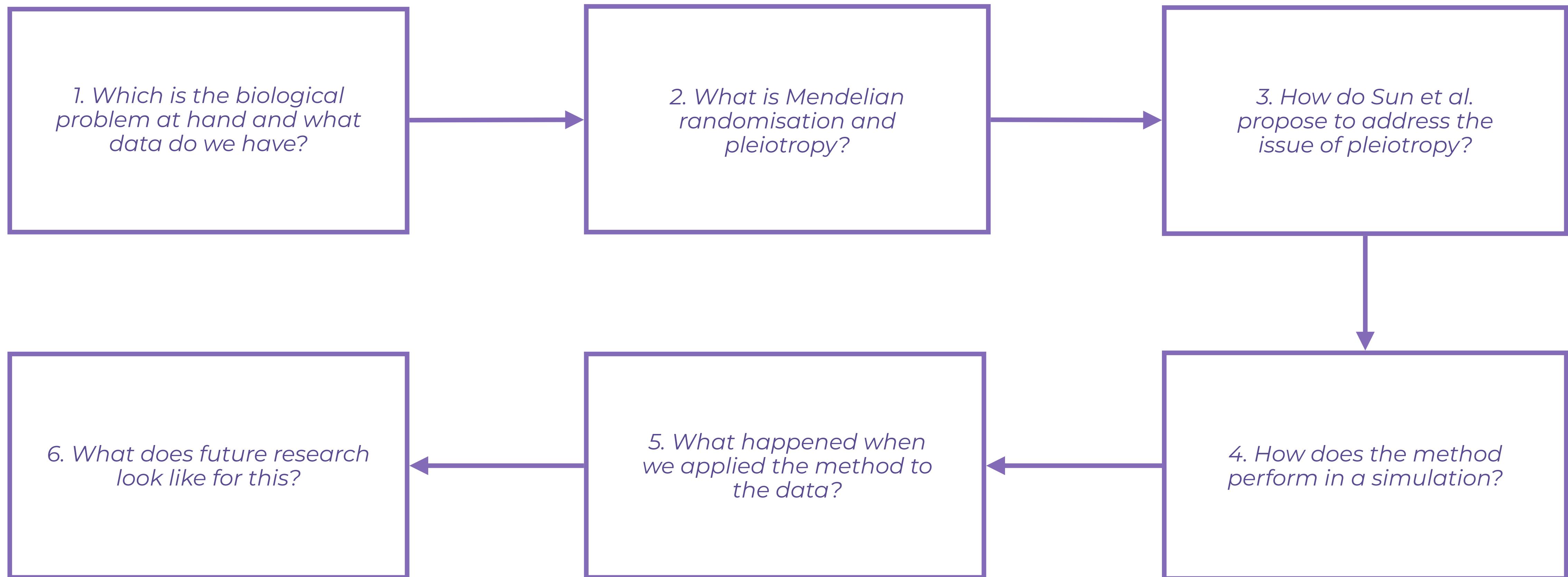
Why do we need to talk about pleiotropy?



MAC-MIGS

igc
INSTITUTE OF
GENETICS & CANCER

The pathway for the talk today.

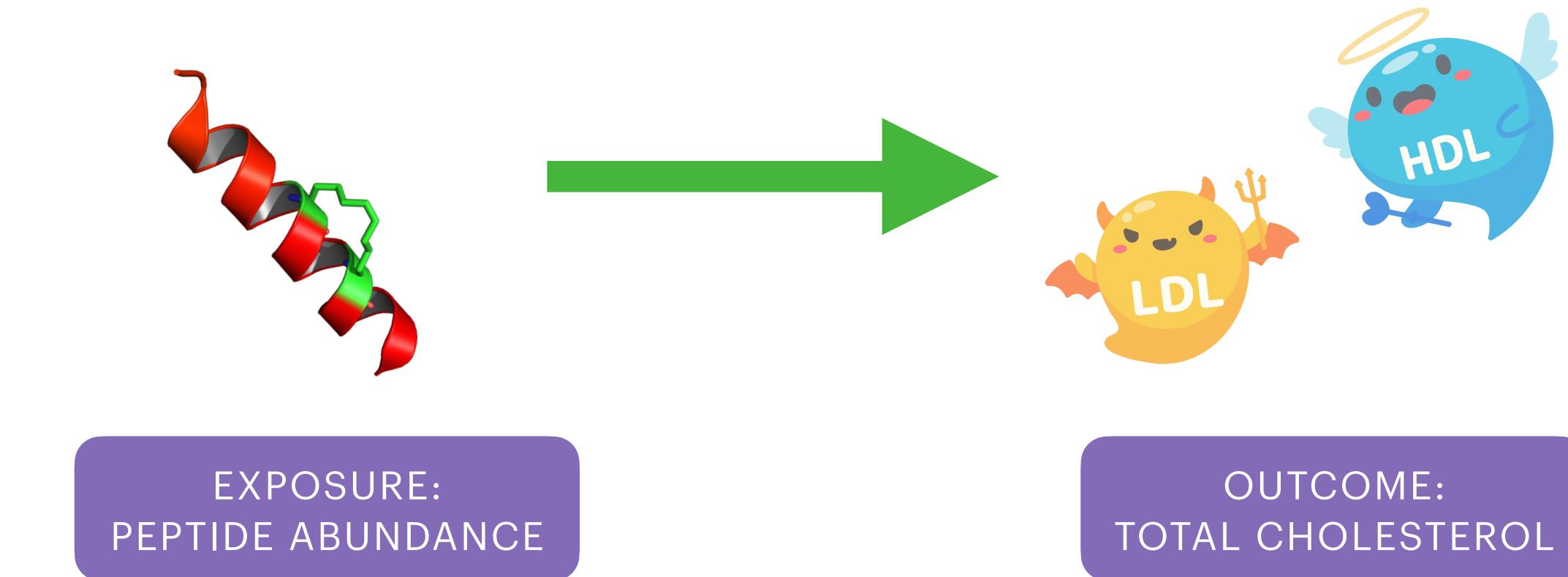


Which is the biological problem at hand and what data do we have?

1. Which is the biological problem at hand and what data do we have?

Which is the biological problem at hand and what data do we have?

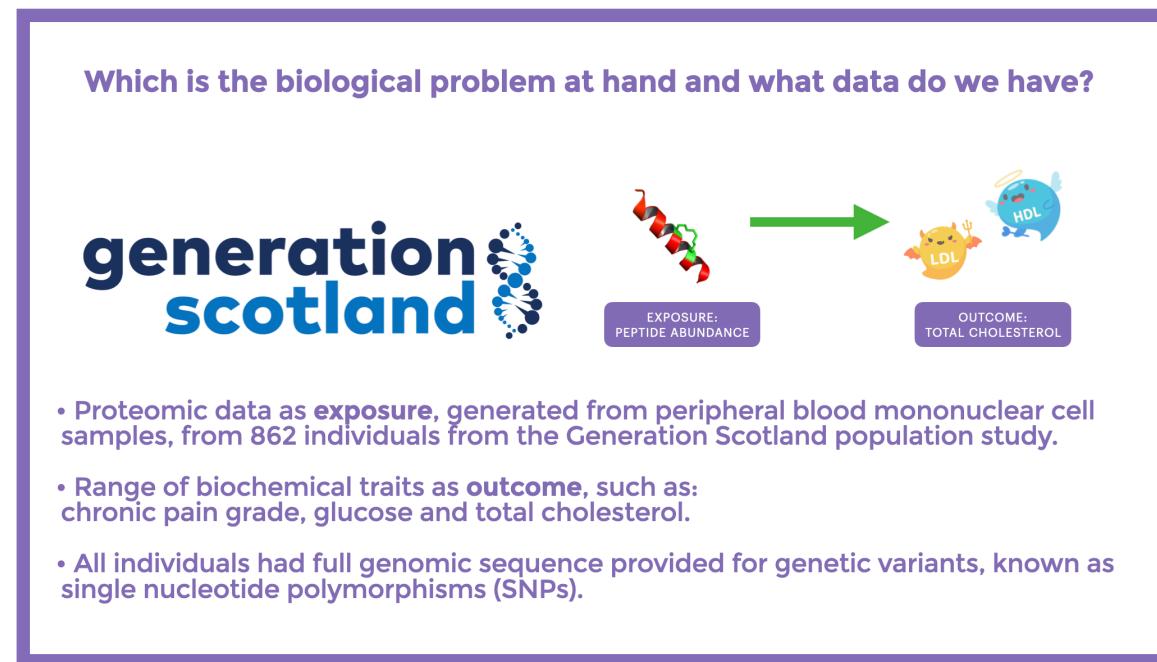
**generation
scotland** 



- Proteomic data as **exposure**, generated from peripheral blood mononuclear cell samples, from 862 individuals from the Generation Scotland population study.
- Range of biochemical traits as **outcome**, such as: chronic pain grade, glucose and total cholesterol.
- All individuals had full genomic sequence provided for genetic variants, known as single nucleotide polymorphisms (SNPs).

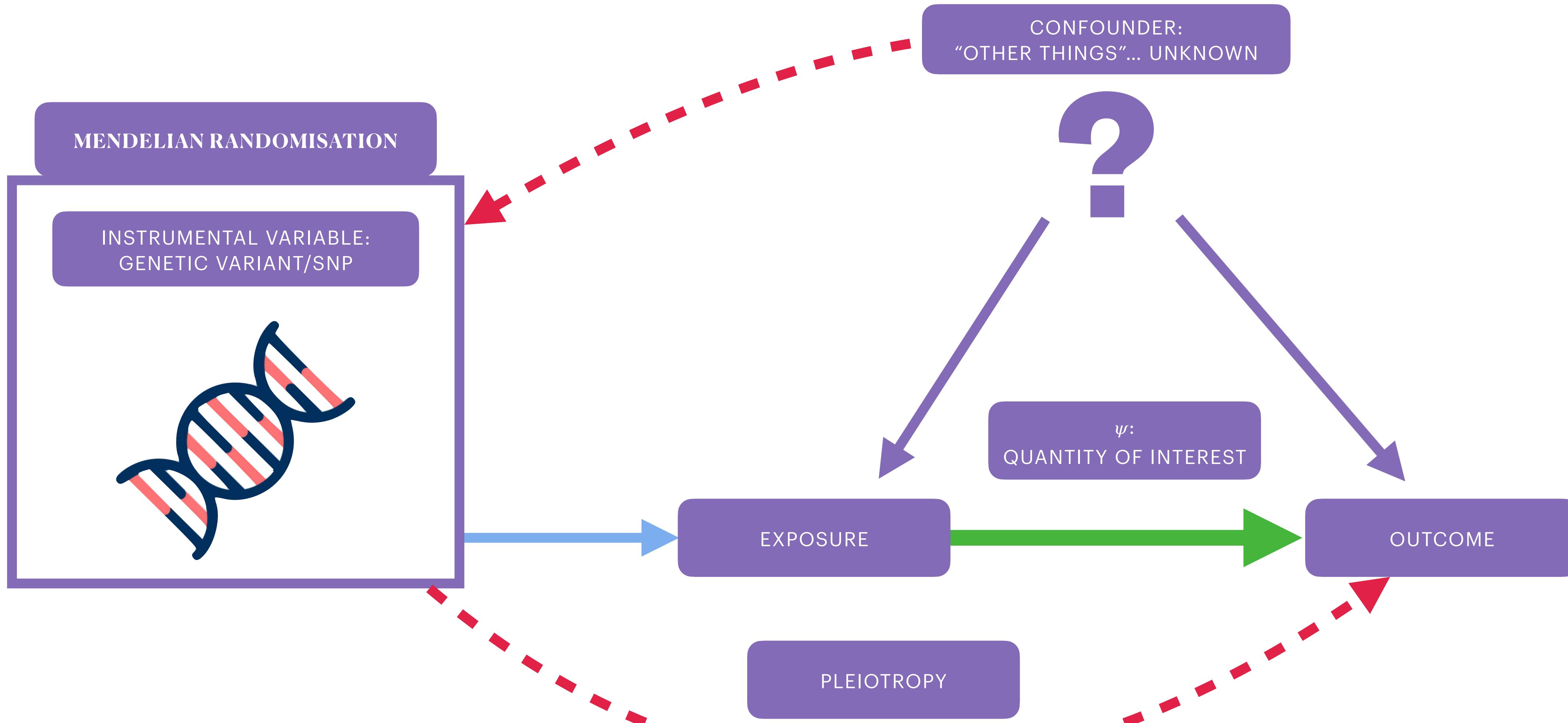
What is Mendelian randomisation and pleiotropy?

1. Which is the biological problem at hand and what data do we have?



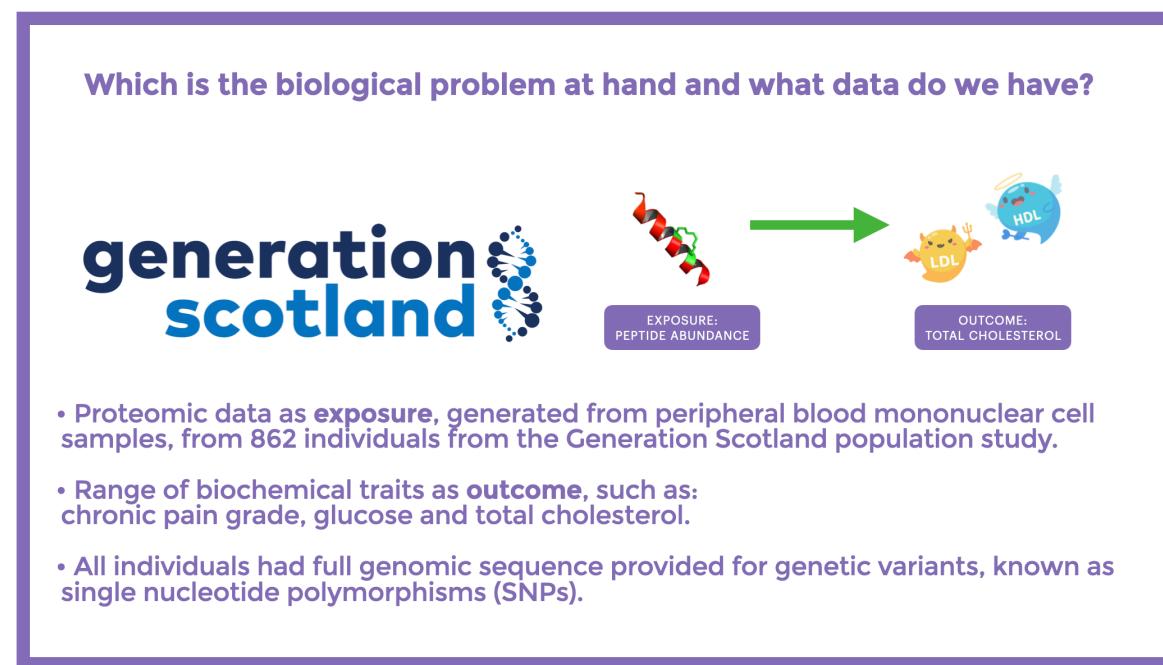
2. What is Mendelian randomisation and pleiotropy?

What is Mendelian randomisation and pleiotropy?

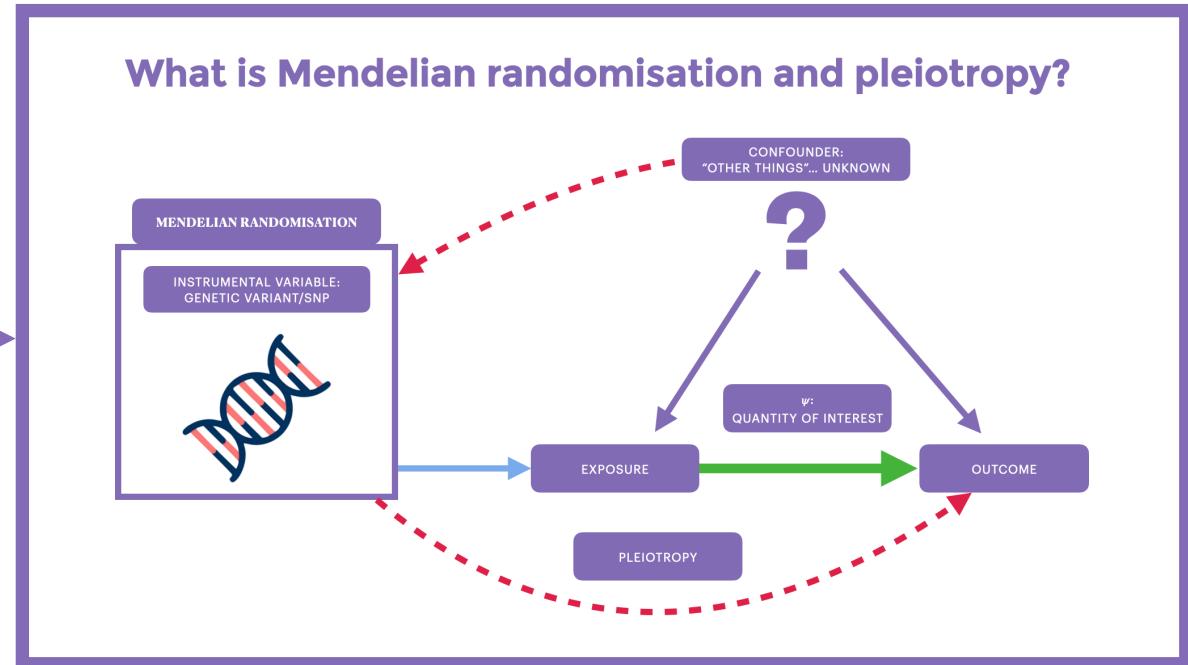


How do Sun et al. propose to address the issue of pleiotropy?

1. Which is the biological problem at hand and what data do we have?



2. What is Mendelian randomisation and pleiotropy?



3. How do Sun et al. propose to address the issue of pleiotropy?

How do Sun et al. propose to address the issue of pleiotropy?

G-estimation for ex ante invalid instrumental variables.

$$\mathbb{E}[D(Z)(Y - \beta^* A)] = 0$$

$$D_\gamma(Z; \mu^*) = \left\{ \prod_{s \in \alpha(1)} (Z_s - \mu_s^*), \dots, \prod_{s \in \alpha(d_\gamma)} (Z_s - \mu_s^*) \right\}^T.$$

$\mathcal{O} = (Z, A, Y)$

Z – Instrumental Variable

A – Exposure

Y – Outcome

γ – Number of assumed valid IVs

K – Total number of IVs

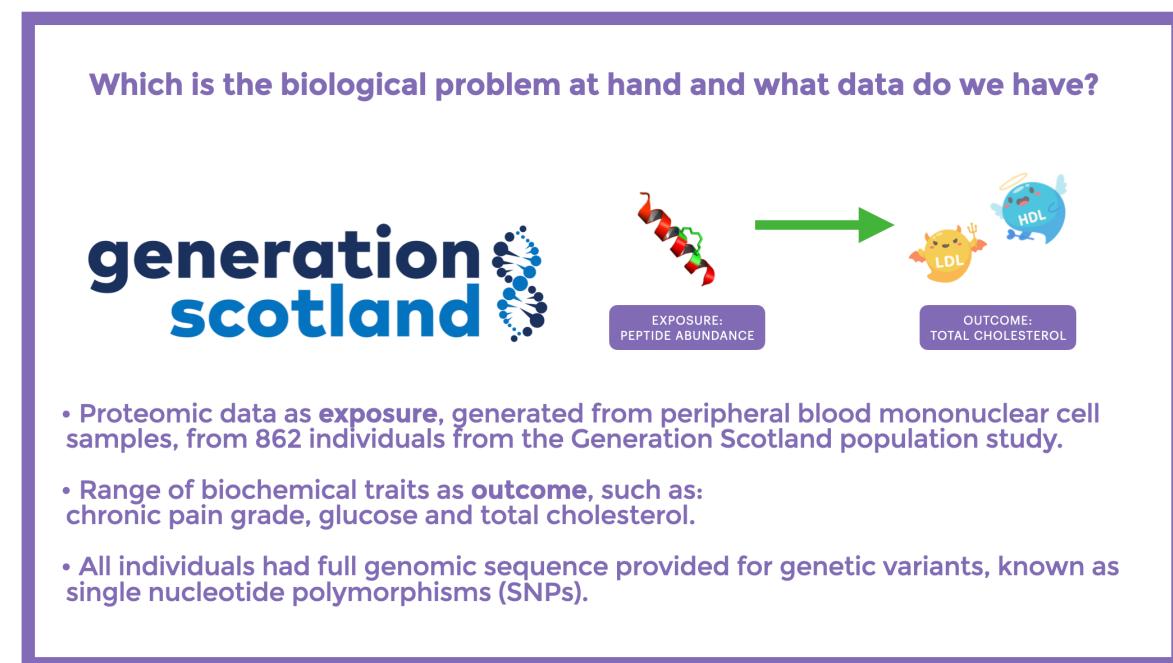
$\alpha \in \{l \subseteq \{1, \dots, K\} \mid K - \gamma + 1 \leq |l| \leq K\}$

$\mu_i^* = \mathbb{E}[Z_i]$

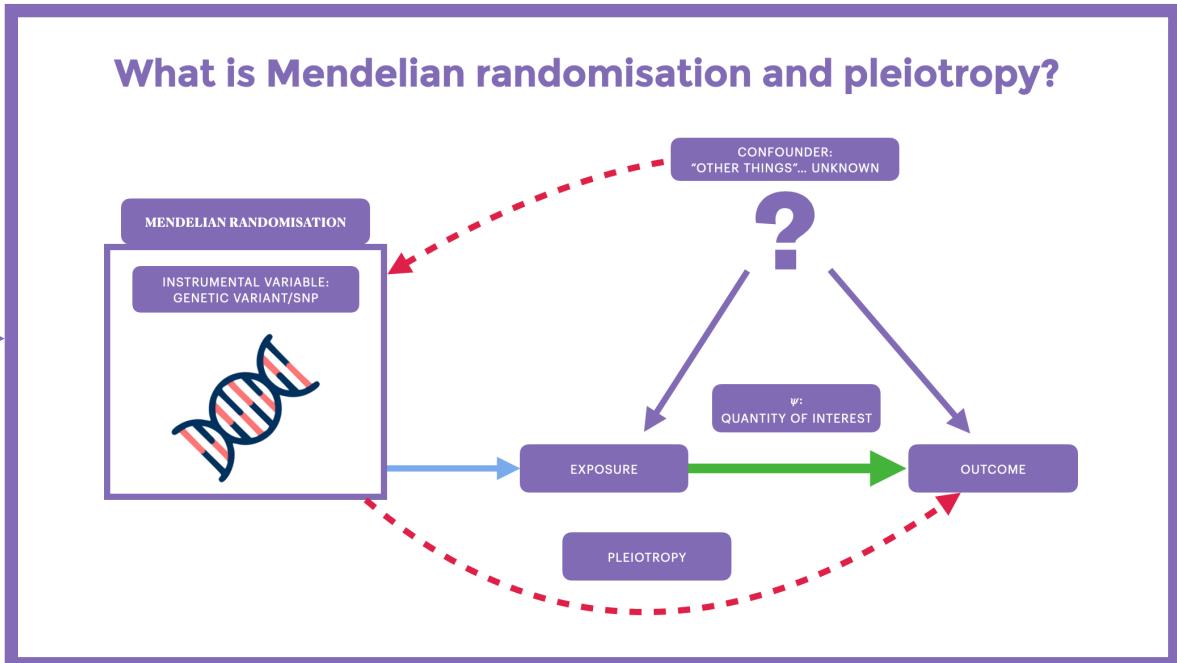
- Best practice to report results for all gamma, as to reflect uncertainty in the estimate.

How does the method perform in a simulation?

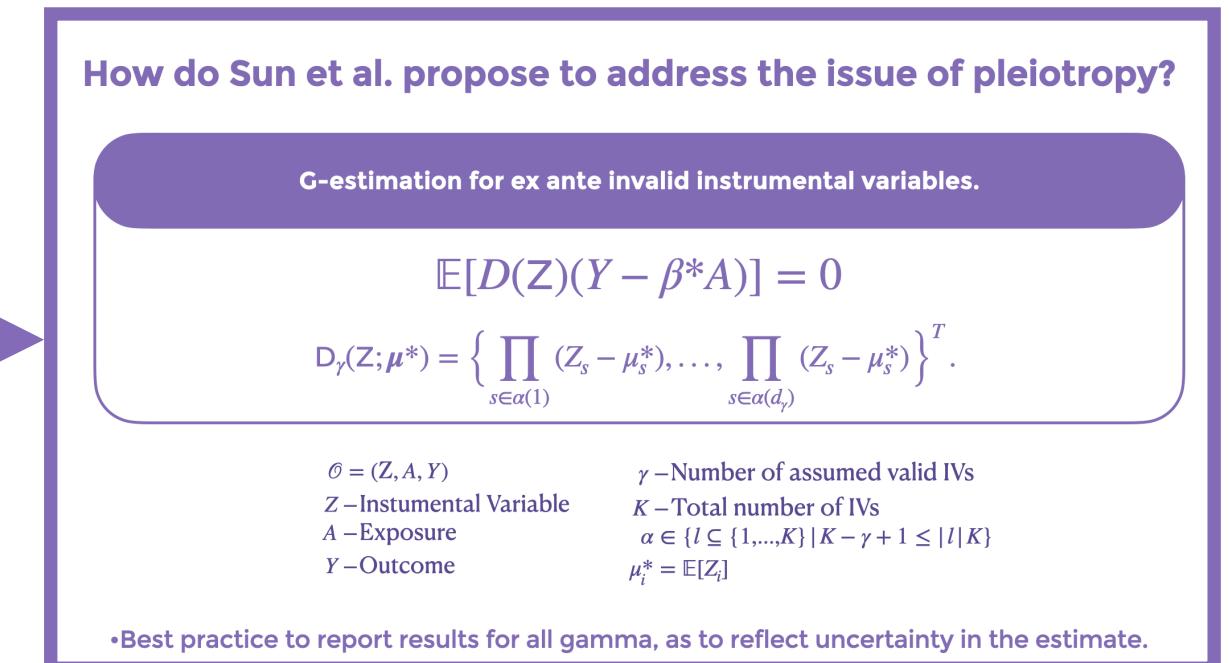
1. Which is the biological problem at hand and what data do we have?



2. What is Mendelian randomisation and pleiotropy?



3. How do Sun et al. propose to address the issue of pleiotropy?

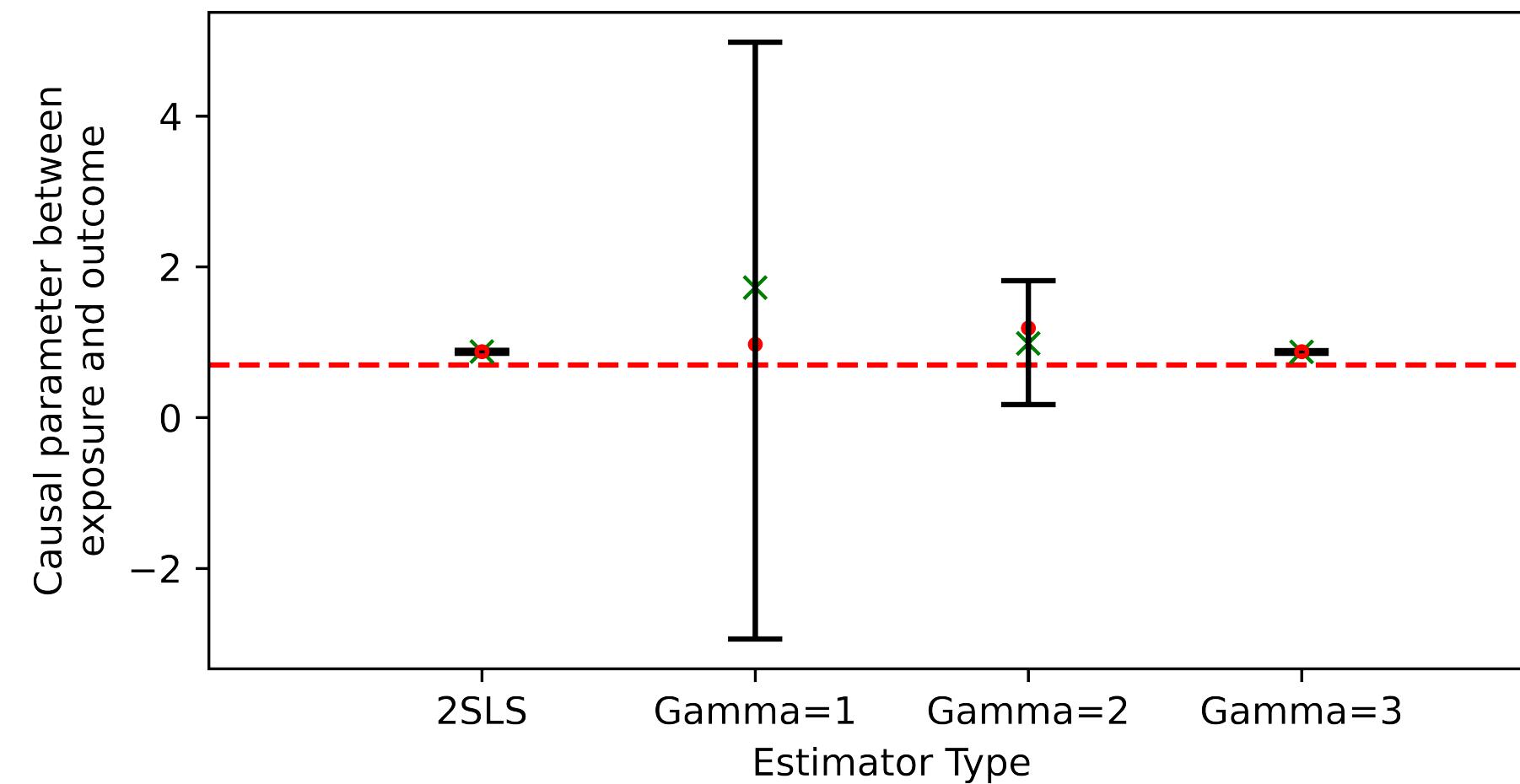


4. How does the method perform in a simulation?

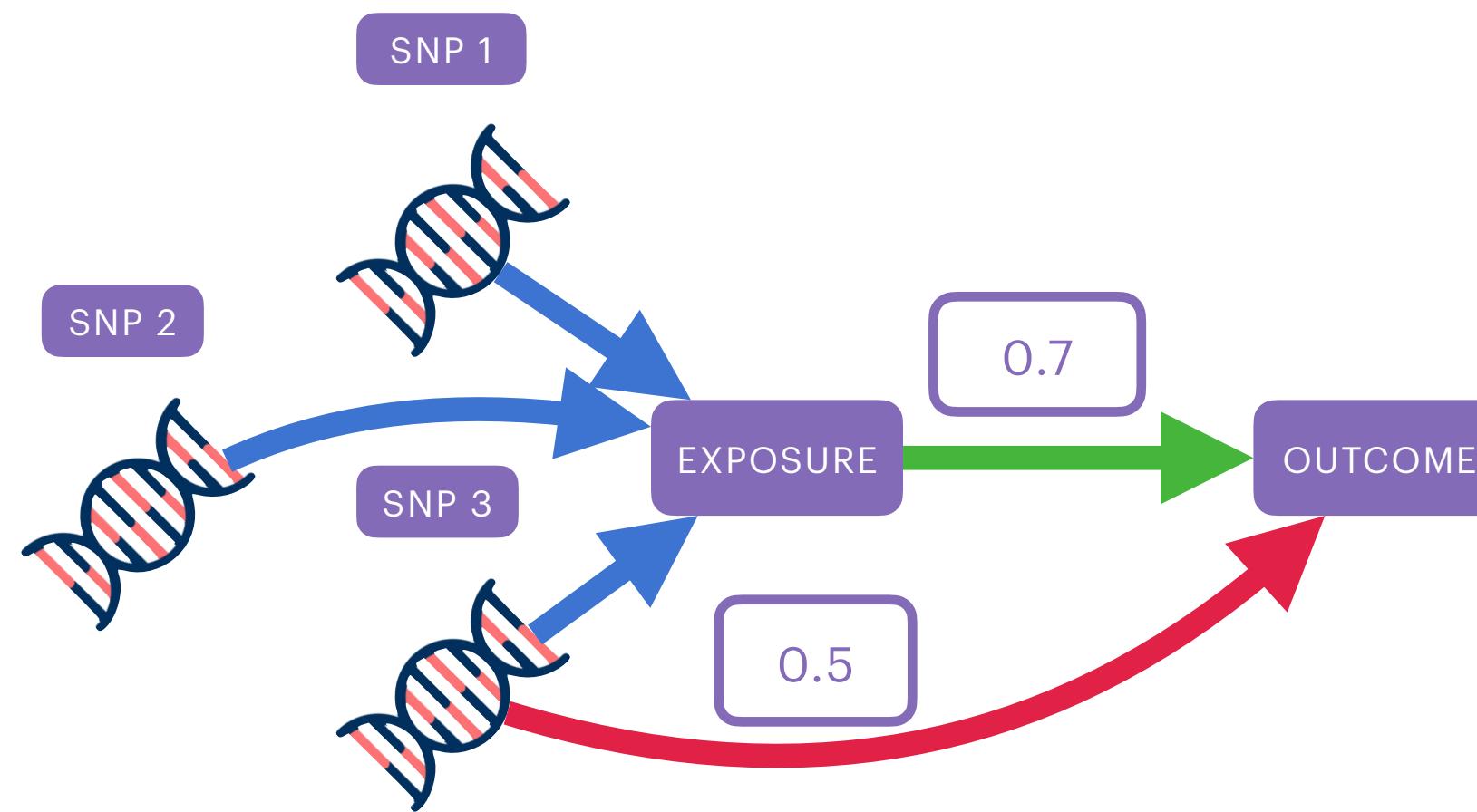
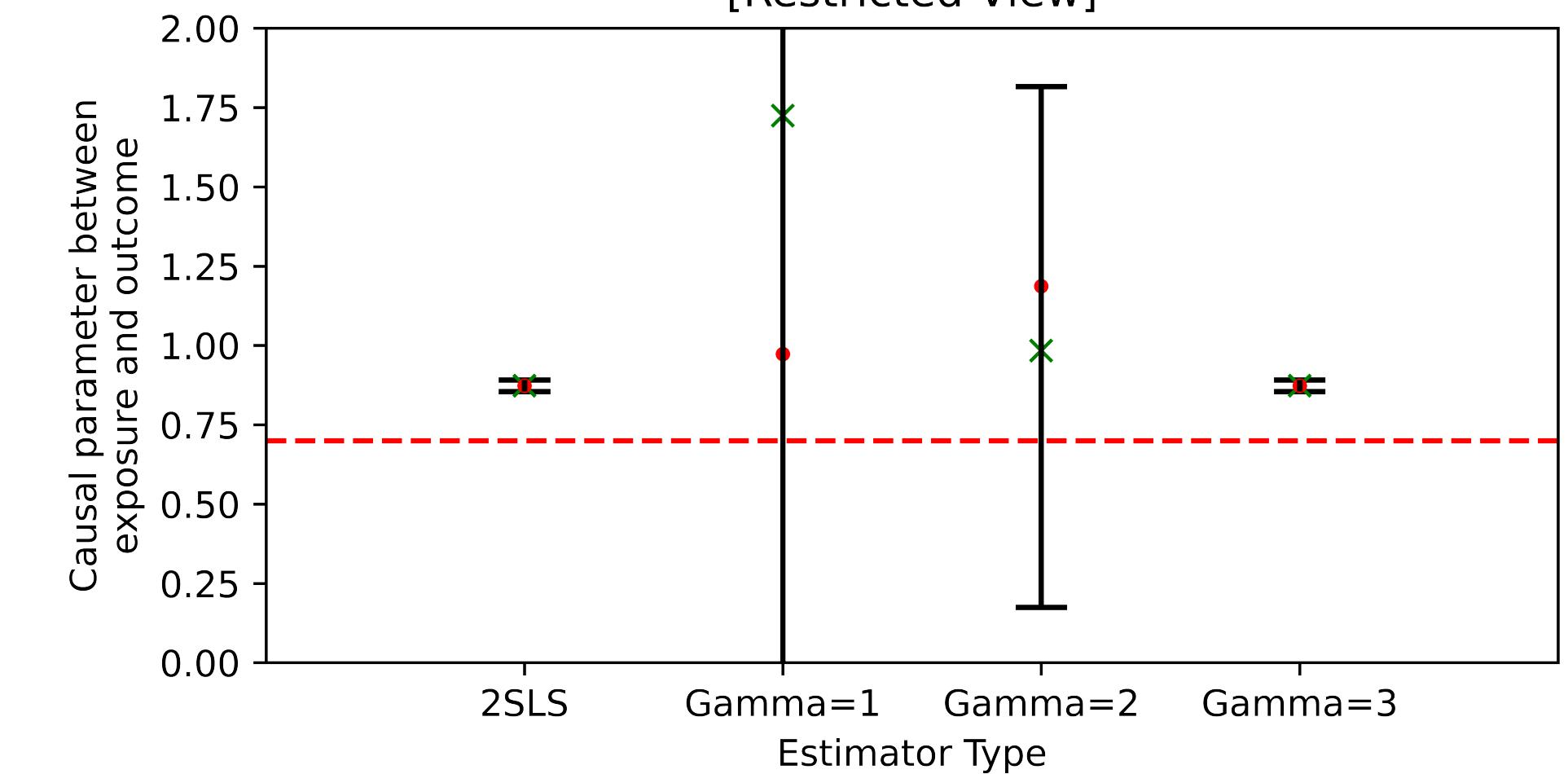
How does the method perform in a simulation?

Note: The green crosses (x) on the plots annotate the mean of the bootstrap runs, whilst the red dot (.) is the causal estimate.

Plot showing different estimator performance for estimating the causal parameter between exposure and outcome.
[Restricted View]



Plot showing different estimator performance for estimating the causal parameter between exposure and outcome.
[Restricted View]



Here we have the data generated by the following process:

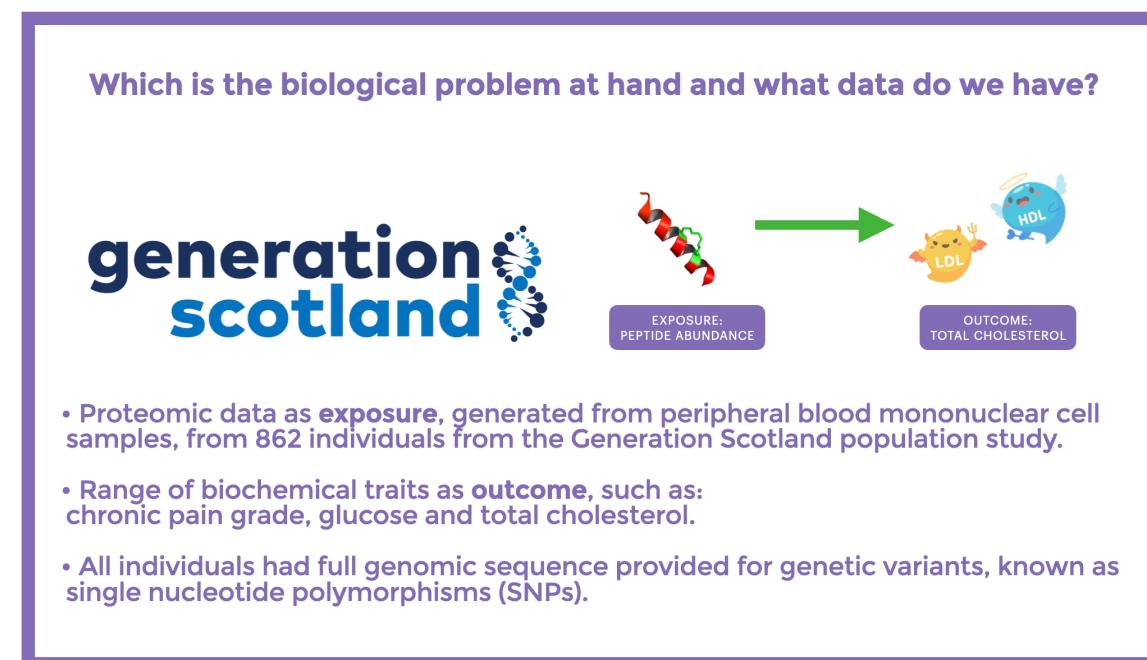
$$\begin{aligned}Y &= 0.7 \cdot A + \Pi \cdot Z + \epsilon_1 \\A &= Z_1 + Z_2 + \epsilon_2 \\\Pi &= (0, 0, 0.5)\end{aligned}$$

$$Z_i \sim \text{Bin}(2, 0.3)$$

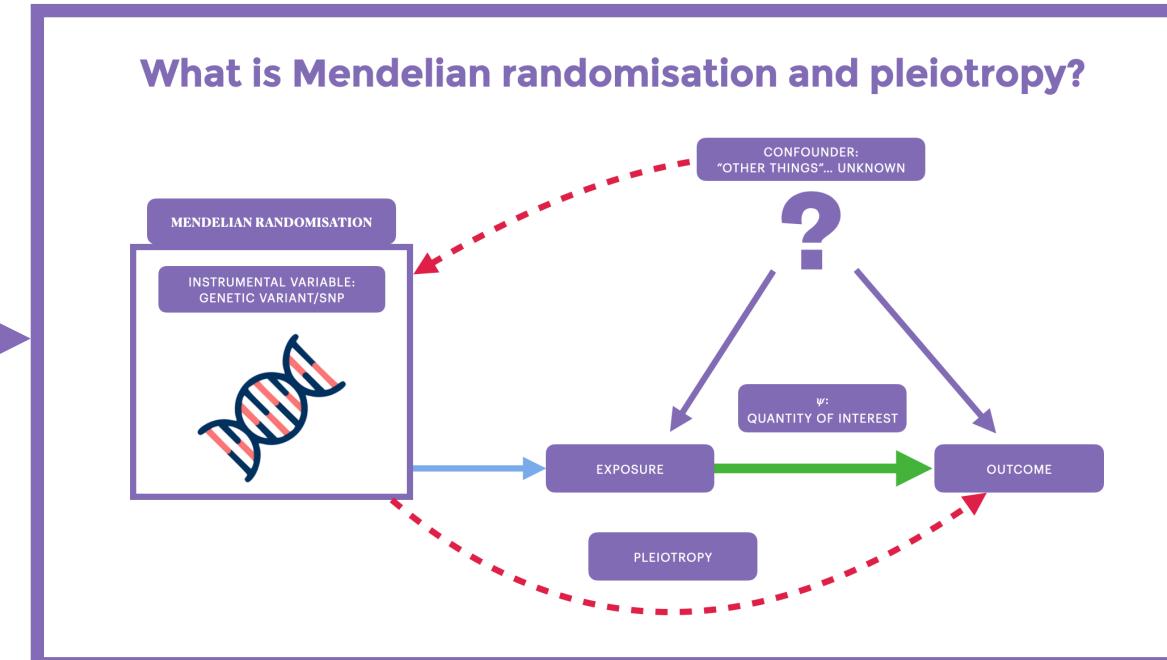
$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim \mathcal{N} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix} \right\}$$

What happened when we applied the method to the data?

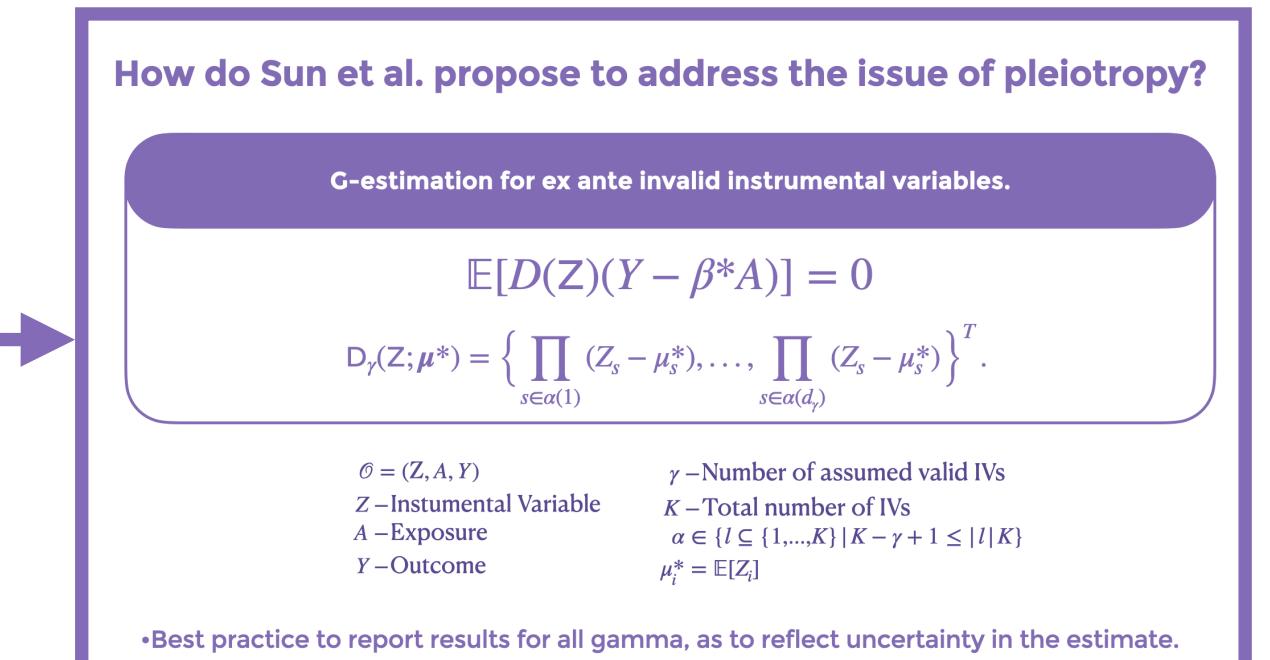
1. Which is the biological problem at hand and what data do we have?



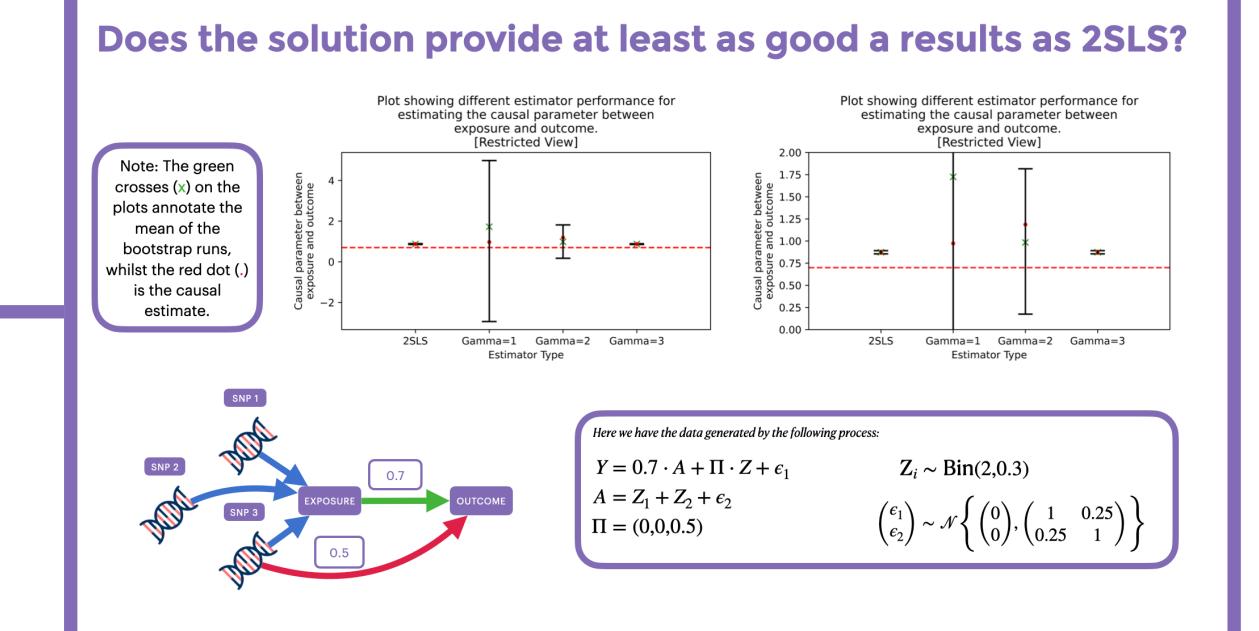
2. What is Mendelian randomisation and pleiotropy?



3. How do Sun et al. propose to address the issue of pleiotropy?

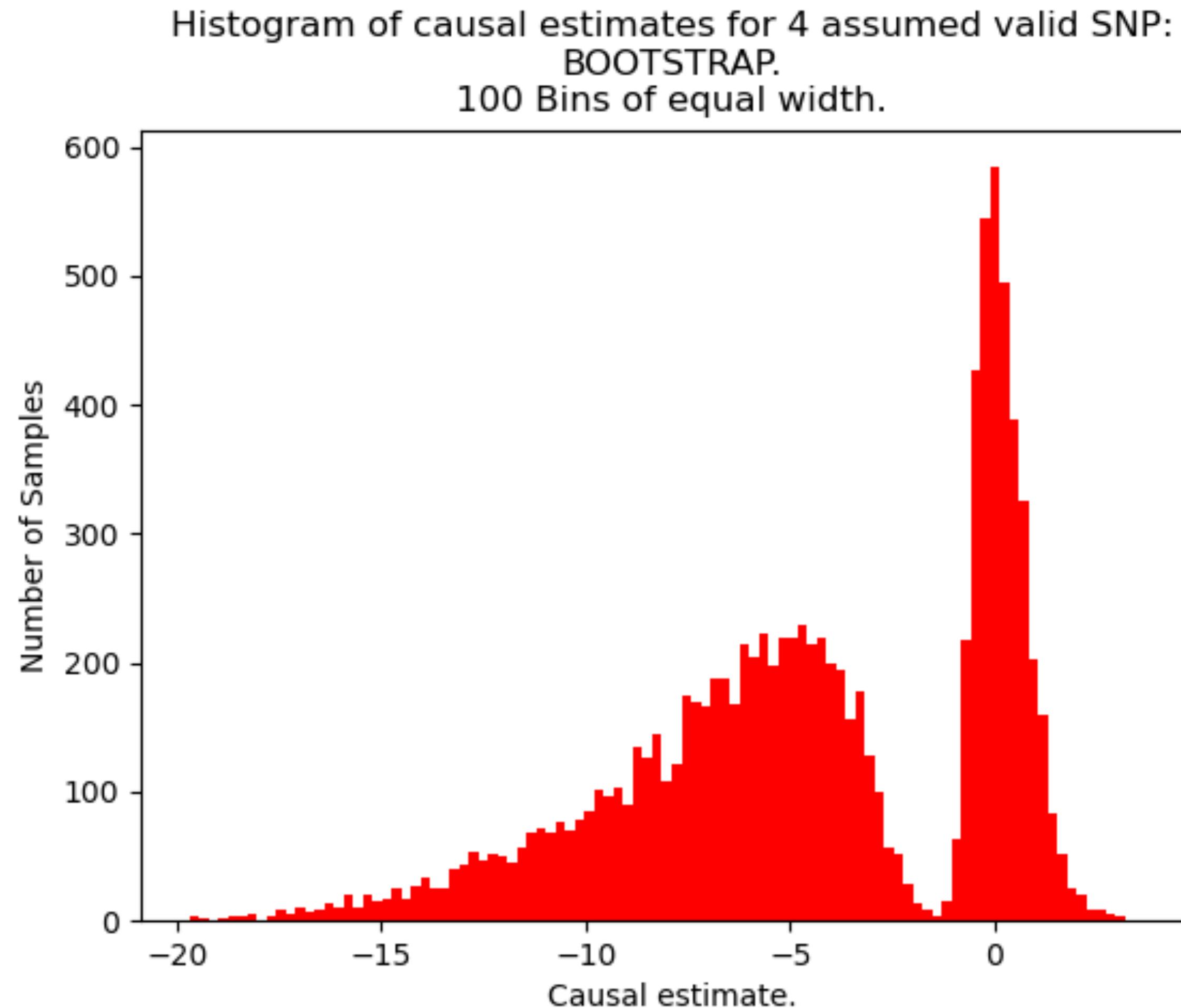


5. What happened when we applied the method to the data?



4. How does the method perform in a simulation?

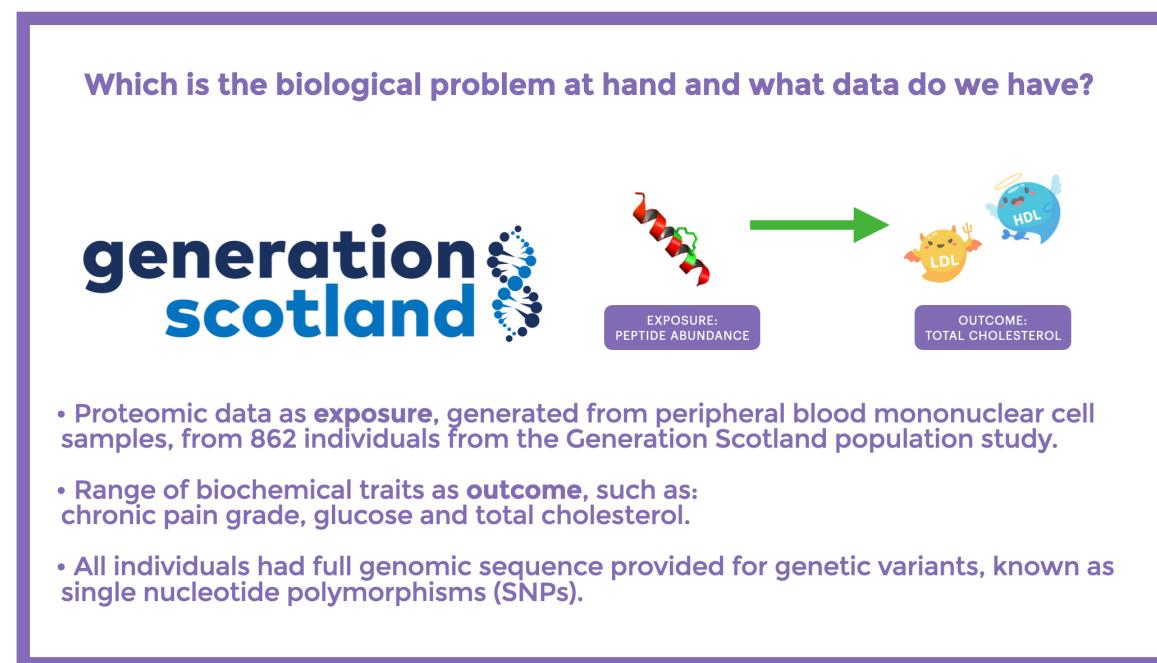
What happened when we applied the method to the data?



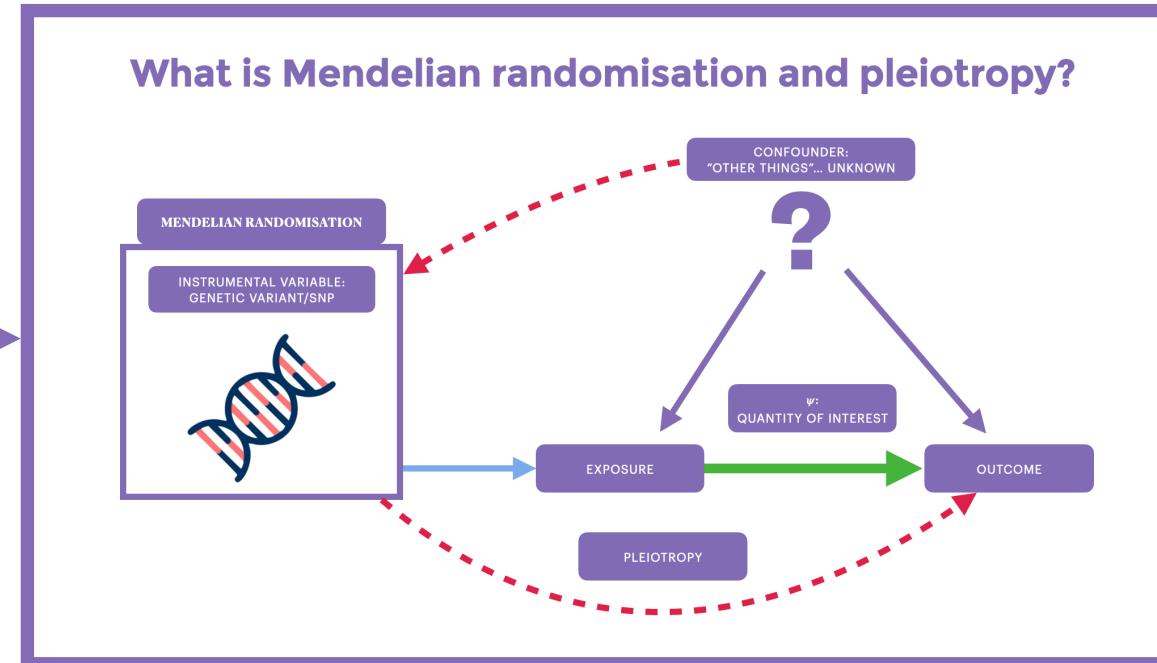
- Multimodal bootstrap of causal estimate.
- Resulted as a consequence of a strong instrument (SNP to peptide association) becoming weak when the outcome data (chronic pain grade) was introduced.
- In this case, the sample size was reduced from 862 to 385.

What does future research look like for this?

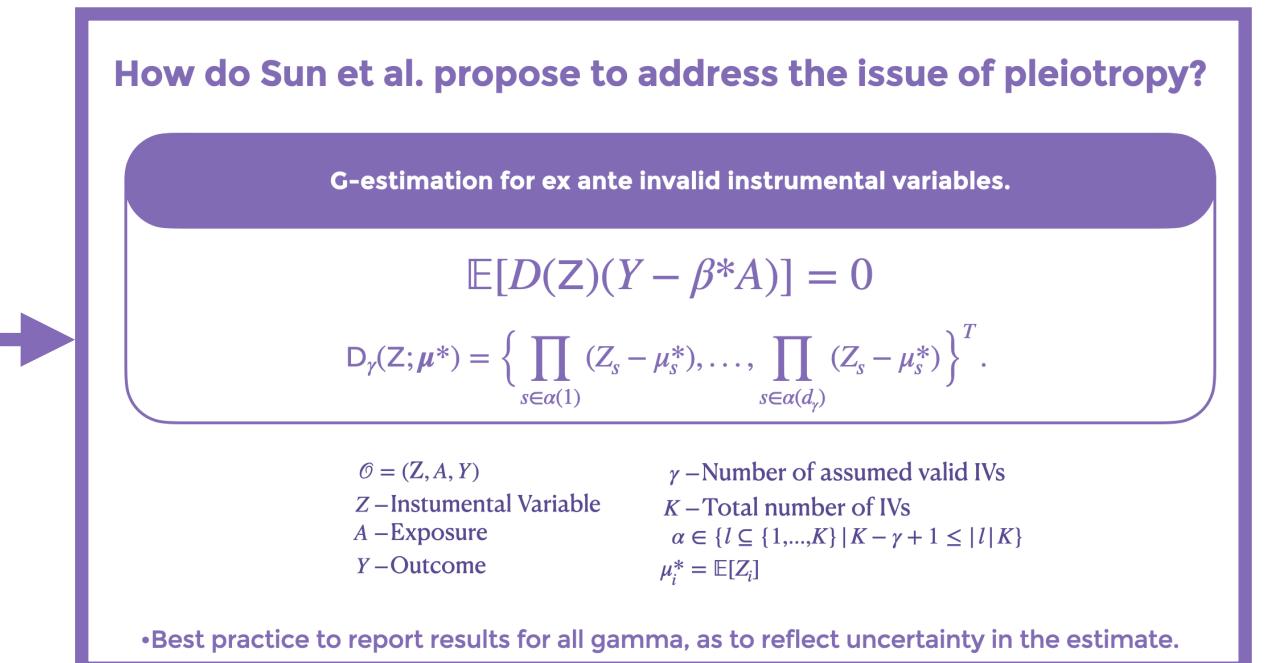
1. Which is the biological problem at hand and what data do we have?



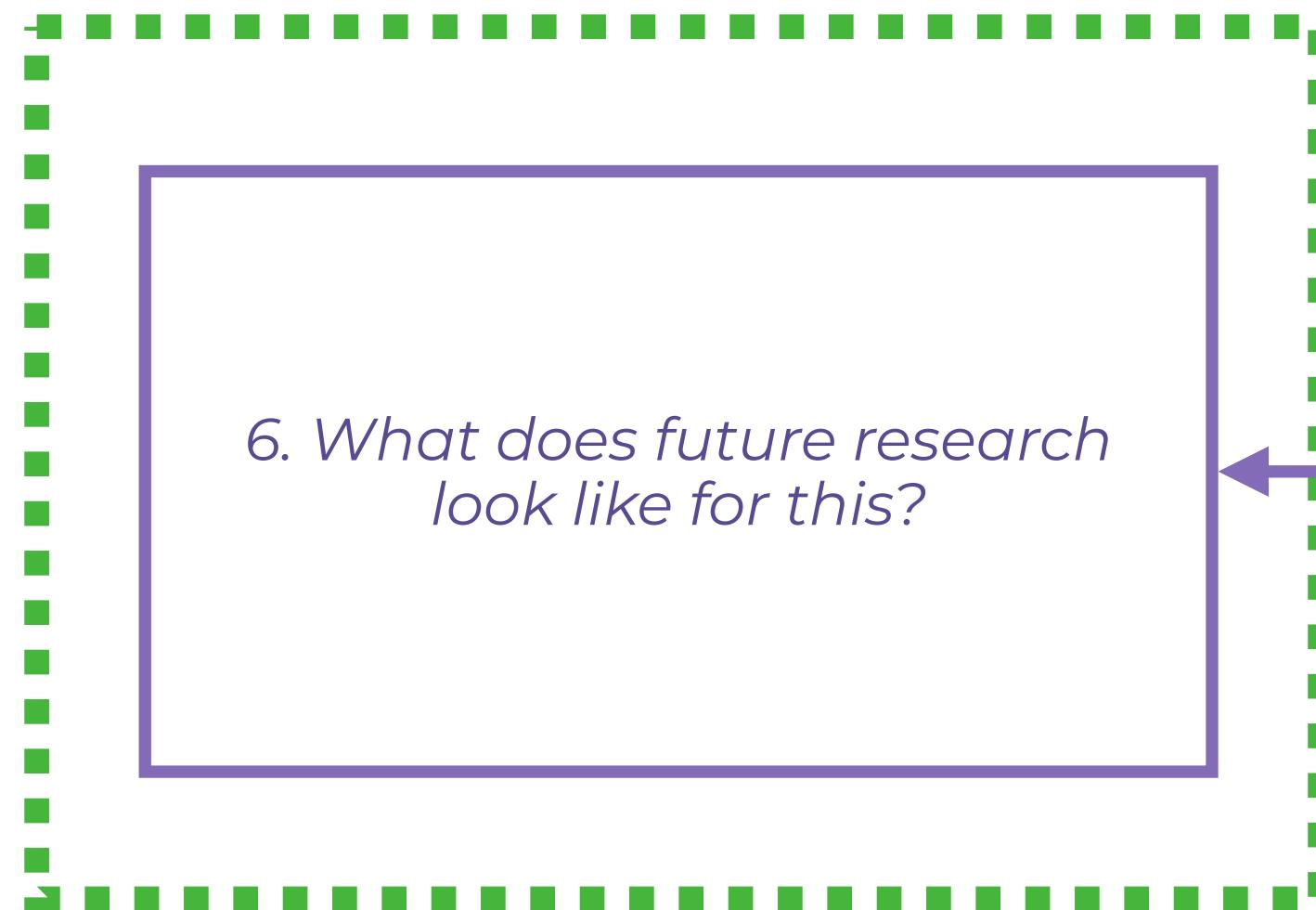
2. What is Mendelian randomisation and pleiotropy?



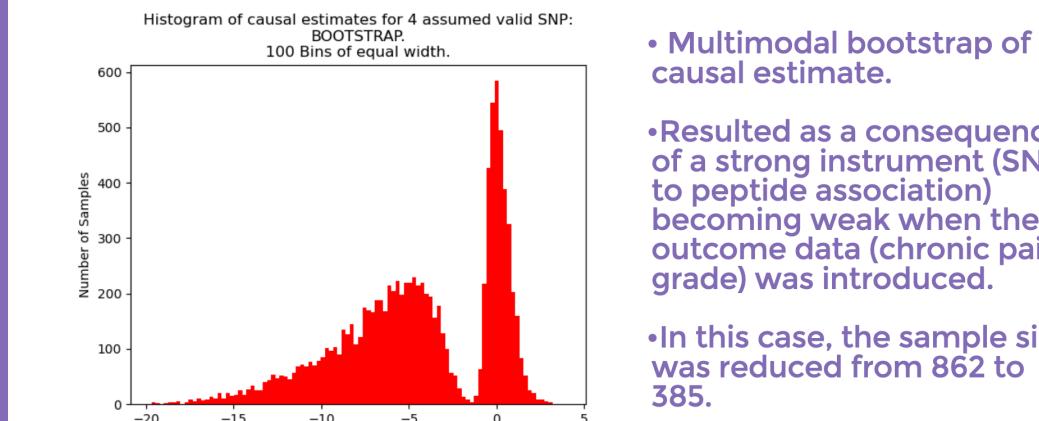
3. How do Sun et al. propose to address the issue of pleiotropy?



6. What does future research look like for this?

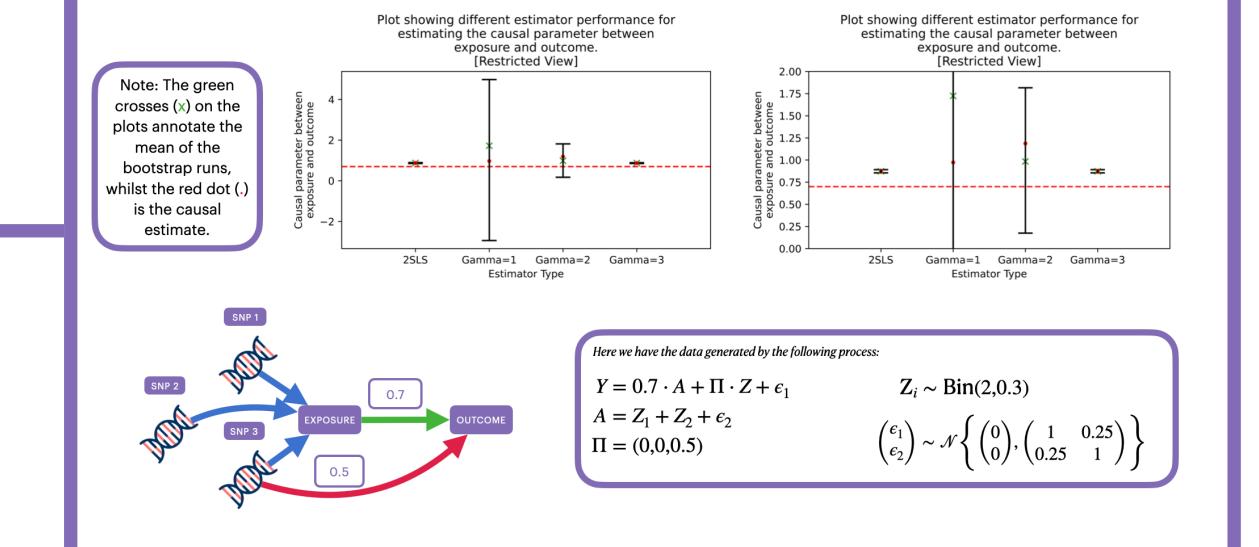


Why did we have issues with the results?



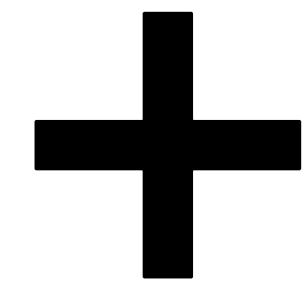
5. What happened when we applied the method to the data?

Does the solution provide at least as good a results as 2SLS?

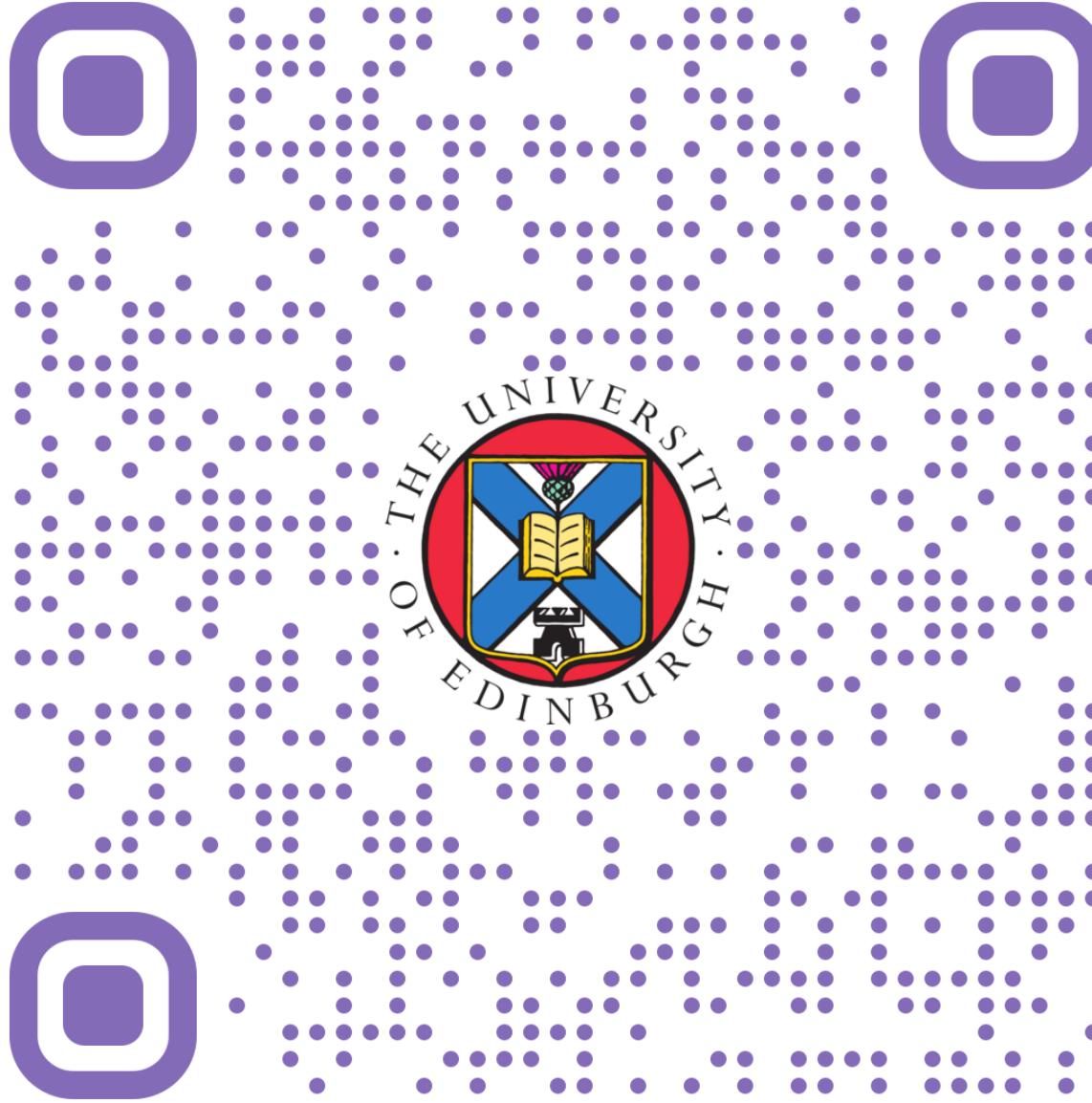


4. How does the method perform in a simulation?

What does future research look like for this?



Summary results from Generation Scotland verified as pleiotropically robust (or not!) by specific individual protein testing in the UK Biobank.



Beentjes-Khamseh-Ponting Groups

Dr Sjoerd Beentjes

Dr Andrew Bretherick

Dr Ava Khamseh

Prof Chris Ponting



PhD Funding:

EPSRC Centre for Doctoral Training in Mathematical Modelling, Analysis and Computation (MAC-MIGS) funded by the UK Engineering and Physical Sciences Research Council (grant EP/S023291/1), Heriot-Watt University and the University of Edinburgh

Generation Scotland Ethics Statement:

Generation Scotland participants provided written informed consent. Ethical approval was provided by the East of Scotland Research Ethics Service committee on research ethics (REC references 15/ES/0040).

With thanks to those participants.

UK Biobank Access:

Research and usage of UK Biobank data will be conducted under the approved UK Biobank Application 91924.

Paper by Sun et al.

B Sun, Z Liu, E J Tchetgen Tchetgen, Semiparametric efficient G-estimation with invalid instrumental variables, *Biometrika*, 2023; doi.org/10.1093/biomet/asad011