Causal effect variations across geographic regions

In this PhD project, we will explore variations of causal effects **across geographic regions**. In standard Mendelian Randomisation (MR), one estimates an ***constant causal effect*** between a risk factor and an outcome, using genetic markers as instrumental variables. In reality this effect may not be homogeneous in a population (*1, 2*), neither in space or time (*3*), nor across the full range of exposure values (leading to non-linear effects) (*4*). Thus, ***in different contexts the same exposure may have different impact on disease*** predisposition.

Specifically, we will interrogate ***environment-specific causal effects*** by exploring the variability of ***causal effects across geographic regions in the UK*** (and ***between countries*** – see below). For this, we model the exposure (), the outcome () and a given genetic instrument () in region , where individuals are observed. The respective trait values for these three variables in region are denoted by vectors and . We assume that regions are uncorrelated in terms of environment and noise and within a region SNP effects and causal effects are homogeneous. Thus, ***the model for region can be written*** as

and

where is the region-specific genetic effect of on and is the region-specific causal effect of on . Since the errors and are correlated (due to potential confounders), the causal effect cannot be directly estimated. Instrumenting with (as long as the instrument satisfies the MR assumptions) allows the estimation of the region-specific causal effects:

The remaining part of these traits and include correlated (due to confounding factors) errors with the following distribution

The region-specific effects are modelled to be distributed as follows

where is the overall genetic effect (-on-) across regions and is the overall causal effect of on . The key estimated parameters are and . The first parameter reveals the extent of region-specific genetic effects acting on , the ***second parameter assesses the heterogeneity of the causal effect of on*** . A significant estimate for would translate into region specific causal effects.

Region-specific causal effects can be explored at a much lower resolution, namely if the ***causal effects within genetically more distant populations are compared***. Hence, we propose to additionally perform ***Mendelian Randomisation in various population (ancestry) groups***. The effect size of a SNP may vary in different populations due to discrepancies in MAF and LD between the interrogated variant and the common causal variant in the different ancestry groups(*5, 6*). In such situations, however, the LD difference cancels out in the ratio estimates, hence would not lead to biased MR results (only to inflated variances). The second scenario is the presence of population specific causal effect size differences, but again these cancel out in MR analysis, unless there are population-specific pleiotropic effects. In summary, ***performing 2-sample MR using summary statistics from different populations is likely to be unbiased by LD-, allele frequency- or effect size differences between populations***. Summary statistics are readily available from the UK Biobank

(http://www.nealelab.is/uk-biobank), Finnish Biobank (https://r8.finngen.fi/), the ***Estonian*** Biobank (application pending) and the Biobank Japan (https://pheweb.jp/). Future releases are expected from the All of Us study(*7*) to complement with association results from individuals with African or Hispanic ancestry.

References

1. J. Sulc *et al.*, Composite trait Mendelian randomization reveals distinct metabolic and lifestyle consequences of differences in body shape. *Commun Biol* **4**, 1064 (2021).
2. J. Sulc, T. W. Winkler, I. M. Heid, Z. Kutalik, Heterogeneity in Obesity: Genetic Basis and Metabolic Consequences. *Curr Diab Rep* **20**, 1 (2020).
3. T. G. Richardson, E. Sanderson, B. Elsworth, K. Tilling, G. Davey Smith, Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. *BMJ* **369**, m1203 (2020).
4. J. D. Angrist, G. W. Imbens, 2-Stage Least-Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *J Am Stat Assoc* **90**, 431-442 (1995).
5. Y. Wang *et al.*, Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat Commun* **11**, 3865 (2020).
6. R. Mester *et al.*, Impact of cross-ancestry genetic architecture on GWAS in admixed populations. *bioRxiv*, 2023.2001.2020.524946 (2023).
7. A. Investigators, The “All of Us” research program. *New England Journal of Medicine* **381**, 668-676 (2019).