**The causal map of 150 complex traits and diseases: a first draft**

Gibran Hemani, Philip Haycock, Jie Zheng, Tom Gaunt, George Davey Smith

We have complete summary data for 150 complex traits and diseases, obtained from GWAS performed on large samples of Europeans. It is now somewhat trivial to use MR-Base to obtain 2-sample Mendelian randomisation (2SMR) estimates of each trait against every other trait. But this raises two fundamental questions: 1) How can we navigate through the myriad of available 2SMR methods to obtain the most reliable causal effect estimate for each relationship; and 2) What is the utility of such a matrix of effects?

We performed extensive simulations to explore how the reliability of a causal effect estimate between two traits can be improved if we have data on many other traits. We demonstrate that when GWAS statistical power is high, and traits are related in complex networks, it is very likely that some instruments will be invalid because they are primarily acting on confounders or are having a reverse causal effect. Using the Rucker framework to account for horizontal pleiotropy does not fully avoid this problem, nor do median or mode based estimators. Using an iterative approach to navigate through the network of traits to update pairwise estimates by omitting instruments that are likely to be unreliable does substantially reduce bias and false discovery rates, and can often improve power by reducing heterogeneity.

We used the approach to construct a first draft of the map of causal relationships between 150 complex traits and diseases and after Bonferroni correction for multiple testing we identify many hundreds of putative associations. We demonstrate that construction of this matrix can substantially improve power in MR, because the number of putative exposure-outcome associations increases dramatically when allowing for causal chains to include multiple intermediate traits. We also show that subsets of the matrix can be orthogonalised to decompose the total effect estimates into direct and indirect relationships for an arbitrary number of traits. However, the empirical results from this analysis make it clear that apparently robust MR estimates can be obtained for biologically impossible trait relationships, and the automation of causal inference must be approached with caution.