

Understanding the genetic and causal relationship between common health outcomes and mortality.

Ole Sahlholdt Hansen

Student number: 201708518

Supervisor: Bjarni Jóhann Vilhjálmsson



Bioinformatics Research Centre

Aarhus University

10th of June 2022

Table of Contents

Abstract	2
Introduction	3
Background	3
Introduction to the project	3
GWAS	5
LD score and LD score regression (LDSC)	7
Mendelian Randomisation	11
GSMR	13
LCV	15
Methodology	16
Correlation analysis using LDSC	17
Causal inference using MR	18
LCV	19
Results	20
LDSC results	20
GSMR	21
LCV	26
Discussion	27
Correlation analysis	27
GSMR analysis	27
LCV analysis	30
Input data and bias	31
Concluding remarks	33
Acknowledgements	33
References	34
Appendix: Source code and supplementary figures	42

Abstract

Inferring true causal relationships in epidemiology is notoriously challenging. Still, the benefits that can be reaped from finding methods that can infer such associations are vast (Smith and Hemani, 2014). Indeed, to be able to identify diseases that have a direct causal relationship with an outcome such as mortality, would aid in the development of treatment and prevention of illness.

This thesis concerns finding correlation and causal relationships between common health outcomes and mortality. The LD score regression (LDSC) method was used to compute genetic correlations for diseases such as coronary artery disease, schizophrenia and smoking behaviour etc. with mortality. Additionally, causal inference analysis was performed using the Mendelian Randomisation (MR) software GSMR on the same traits as with LDSC. Additionally, the Latent Causal Variable (LCV) method was used as a comparative measure to GSMR and LDSC. This gave varying results, as each software tended to identify different traits to be in association with mortality. This can have been a consequence of the limitations of the data and confounding.