



Application number/Title: 31063 - Methodological extensions to estimate genetic heritability and shared risk factors for phenotypes of the UK Biobank

Applicant PI: Dr Claire Churchhouse

Applicant institution: Broad Institute, Stanley Center, 415 Main Street, Cambridge MA 02142

Lead Collaborators:

- 1) Dr Alkes Price
- 2) Dr Cecilia Lindgren

Keywords provided by the Applicant PI to describe the research project:

heritability, genetic correlation, epistasis, GWAS, dominance

Application Lay Summary:

1a: We will investigate new approaches to estimating heritability for a wide range of phenotypes and health outcomes, and look at what genetic risk factors are shared between them. We'll extend our existing approach that uses genome-wide association study summary statistics, to consider more sophisticated models that capture a broader range of possible effects (eg. dominance, epistasis). Our new method will simultaneously estimate heritability and the true causal SNP effects that account for LD. We'll gain a clearer understanding of which loci are risk factors for a range of diseases/traits and how these are shared between them.

1b: This work will estimate heritability and detect shared heritability between a large set of phenotypes. In doing so, we aim to find shared genetic risk factors that help explain disease comorbidity and associated risk factors. This work is in line with the UK Biobank's aim of enabling research to improve "prevention, diagnosis and treatment of illness and the promotion of health throughout society". It will not focus on one particular disorder or subtype of disorder, but instead apply these methods to learn about a wide range of traits through a hypothesis-free approach.

1c: We will build on our previously published work to see if more sophisticated models of genetic risk explain more heritability of a broad range of diseases/traits and that which is shared between them. With an approach to infer the true causal effect of genetic variants, rather than that which is confounded by the fact that neighbouring variants often occur together, we obtain a more accurate picture of exactly which ones are protective or pathogenic. This work would involve using imputed genotype data and all available phenotypes (self-reported phenotypes, medical records/registry data).

1d: We are requesting the full UK Biobank cohort, including genetic data on all participants.

We will investigate new approaches to estimating heritability for a wide range of phenotypes and health outcomes, and look at what genetic risk factors are shared between them. We'll extend our existing approach that uses genome-wide association study summary statistics, to consider more sophisticated models that capture a broader range of possible effects (eg. dominance, epistasis). In addition to studying the collected phenotypes, we wish to also study response rates and response characteristics (eg. how often a response is left unanswered) and to examine whether there are any genetic factors that correlate with these response phenotypes. This will allow to better understand response bias in questionnaire studies and thereby inform planning and analysis of epidemiological studies. It may also shed light on the psychometric properties of questionnaires used in UK Biobank. Our new method will simultaneously estimate heritability and the true causal SNP effects that account for LD. We'll gain a clearer understanding of which loci are risk factors for a range of diseases/traits and how these are shared between them.

Extension: We will investigate new approaches to estimating heritability for a wide range of phenotypes and health outcomes, and look at what genetic risk factors are shared between them. We'll extend our existing approach that uses genome-wide association study summary statistics, to consider more sophisticated models that capture a broader range of possible effects (eg. dominance, epistasis). In addition to studying the collected phenotypes, we wish to also study response rates and response characteristics (eg. how often a response is left unanswered) and to examine whether there are any genetic factors that correlate with these response phenotypes. This will allow to better understand response bias in questionnaire studies and thereby inform planning and analysis of epidemiological studies. It may also shed light on the psychometric properties of questionnaires used in UK Biobank. Our new method will simultaneously estimate heritability and the true causal SNP effects that account for LD. We'll gain a clearer understanding of which loci are risk factors for a range of diseases/traits and how these are shared between them.

We wish to extend the scope of our application to include the study of behaviours in response to the UK Biobank questionnaires, and in particular, to examine whether any genetic variants associate with these. Specifically, we propose to look at patterns of response in the questionnaire answers, such as how often a participant chooses not to provide a response, how often they select the most frequent answer, and how often they provide answers that tend to be outliers in the response distributions. We will examine different metrics that capture this information and, as in our original project plan, perform GWAS analysis but now also on these derived characteristics to see if any of these behaviours are, at least somewhat, genetically mediated. Also as in our original scope, we will estimate the genetic correlation between these derived phenotypes and all other traits we have examined to see which, if any, are associated with the response behaviour.