Analysis plan

for GWAS on selected phenotypes at different stages in the lifecourse

Last updated: August 13, 2024

Authors: Grace M. Power, Gibran Hemani, Eleanor Sanderson

Background

Acute, chronic, and recurring, adverse health conditions that emerge in later life are often shaped by processes experienced throughout life. Gaining a better understanding of how exposures at different stages in the lifecourse influence health outcomes is critical to developing more effective disease prevention and treatment strategies. This is of key public health importance.

Lifecourse stratified effects are of primary interest for identifying critical periods in Mendelian randomisation (MR). Inherited genetic variants may have different effects on some exposures at different time periods across the lifecourse (within a population). As a result, we are seeking to establish a consortium dedicated to generating and integrating data using an age-stratified genome-wide association studies (GWAS) approach.

To robustly run MR, valid instrumental variables must be employed which require large-scale datasets comprising phenotype and genotype data. By aggregating data from a wide range of cohorts, we will be able to access larger sample sizes without requiring repeated measures. This approach additionally allows us to remain agnostic about the shape of the Gene-by-Age (GxAge) interaction during the analysis stage and enables us to model it in greater detail at the post-meta-analysis stage.

This consortium will allow us to develop a more comprehensive set of instruments for future MR analyses to be better able to estimate the effects of a range of phenotypes at multiple time periods across the lifecourse on later life outcomes.

Aim

To explore how selected phenotypes at different stages in the lifecourse modify risk, we will:

1. Generate GWAS summary data of age x phenotype across the lifecourse for 24 phenotypes
2. Combine results by meta-analysing at the age x phenotype level across studies
3. Model changes in genetic effects over time at the post-meta-analysis stage
4. Evaluate the influence of biasing mechanisms on GxAge interaction estimates

5. Conduct post GxAge analyses e.g. time-varying MR, and provide a usable tool for researchers interested in using the data generated

Document scope

This Analysis plan will guide you through the Lifecourse GWAS consortium analysis. Here, you will find instructions for data preparation and running code to generate GWASs on time-varying phenotypes. We are collecting data on a comprehensive list of phenotypes (see Supplementary 1 below) every year up until 19 years of age and every five years thereafter.

We have prepared the pipeline to minimise time and energy required by analysts to contribute data to the overall effort, ensure harmonisation across cohorts, and minimise errors. The use of standardised procedures across all samples is critical in order to increase the effectiveness of the subsequent meta-analyses that we be run internally upon receipt of these GWAS. Because there is always a chance of error, we may ask some analyses to be re-run, although we will attempt to keep such requests to a minimum. We encourage analysts to adhere to the data and file organisation structures proposed for the pipeline to facilitate debugging and ease of any subsequent analyses that might be required.

Inclusion criteria

To be included, a cohort will need to provide at least one phenotype for at least one age range. The list of phenotypes is somewhat aspirational and if bandwidth limits the number of phenotypes that can be contributed, that will not be a barrier to contributing where possible. Please discuss which phenotypes to contribute with the core Lifecourse GWAS consortium group if you are uncertain (lifecourse-gwas-group@bristol.ac.uk). More details on the definitions of each phenotype are given in the Phenotype definition section (Supplementary 1).

Overview of analysis required

Below we outline the methodological steps required for generating the GWAS summary statistics for each cohort, summarised in Figure 1. We will provide code for each step in the pipeline. Further details on these steps are given below.

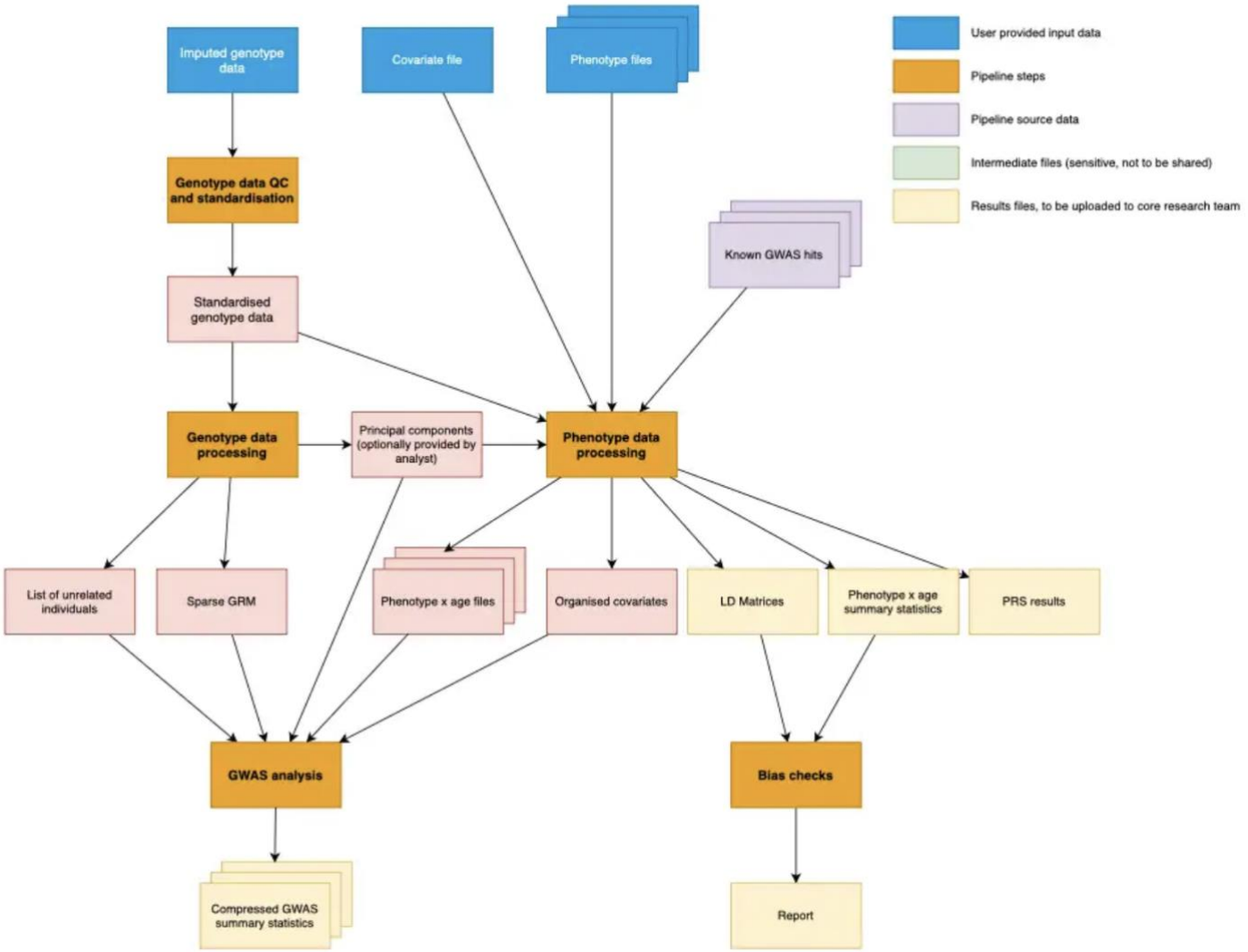

Figure 1. Diagram depicting the steps involved in the pipeline for generating the GWAS summary statistics.

Step 1. Genotype data checks

Genotypes must be imputed to a recent reference panel (TOPMED, HRC, 1000 genomes), provided on build 37/hg19 and with alleles oriented on the forward strand.

Data is expected in plink best guess format, with MAF > 0.01 and imputation INFO > 0.8.

We expect the imputed data has been through rigorous QC, but the pipeline will also perform basic post-imputation quality control (QC).

The pipeline expects a single ancestral group, so if the data can be split into multiple ancestral groups please run the pipeline independently for each ancestral group. For example, clustering individuals in your sample against the 1000 Genomes super populations is one approach to obtain discrete ancestral groups. Guidance will be provided to help with this if needed. Ancestry of course exists on a continuum, and forcing individuals into discrete groups is artificial and arbitrary, but we use this approach to simplify an already complex analytical pipeline.mi

The pipeline will migrate variant identifiers to chr:pos_a1_a2.

Step 2. Genotype data processing

Generate PCs or use user-provided PCs. Check for ancestry outliers. Remove relateds.
For data with substantial relatedness, a sparse GRM will be generated, and PCs will be generated by projecting from unrelateds to relateds.

Step 3. Phenotype data processing

Phenotype summary statistics: Phenotype data is provided by the analyst in long format with each individual's phenotype measure annotated with the age at measurement. In order to ensure comparability of associations across studies we will generate comprehensive summary statistics of phenotype x age distributions, stored without revealing individual level data.

The phenotypes will then be organised using the phenotype file, covariates file, ancestry ID list and PCs per ancestry.

The pipeline will then clean the phenotypes using phenotype-specific transformations and standardisations, and organise the phenotypes and covariates into separate files of phenotype x age categories.

In order to evaluate changing selection over age we will compare LD matrices over time and across studies. The LD matrices will be generated for each phenotype using known GWAS hits from previous biobank or consortia.
Using the same phenotype-associated variants we will also generate PRS for each phenotype and at each time point. PRS-phenotype associations will be generated for each time point.

Step 4. Genome-wide association study

The pipeline will run the GWAS using the sparse GRM, PCs, cleaned genotype, phenotype and covariate files using FastGWA. The pipeline can optionally parallelise this analysis across a cluster.

Resulting GWAS summary statistics will be per phenotype by age. Summary statistics will be compressed into a binary format to reduce storage space requirements.

Uploading results

Each pipeline step will package the logs and shareable results ready for upload. A central server will be provided for analysts to upload the results from each stage. We will encourage analysts to upload the packaged logs and results for inspection by the core research team, so that if there are any issues they can be identified before analysts potentially spend time on subsequent steps.

Contact details

Questions about this analysis plan, or any aspects of the project, can be directed to: lifecourse-gwas-group@bristol.ac.uk

The working group for this consortium consists of:

Grace M. Power grace.power@bristol.ac.uk

Eleanor Sanderson eleanor.sanderson@bristol.ac.uk

Gibran Hemani gibran.hemani@bristol.ac.uk

Genevieve Leyden genevieve.leyden@bristol.ac.uk

David Carslake david.carslake@bristol.ac.uk

Supplementary 1.

Phenotype definitions

Anthropometric

Body Mass Index (BMI)

- Units: kilograms/metres^2 (kg/m2)
- Covariates: sex

Height

- Definition: Standing height
- Units: centimeters (cm)
- Covariates: sex

Waist circumference

- Units: centimeters (cm)
- Covariates: sex

Waist to hip ratio (WHR)

- Definition: Divide the waist measurement (cm) by hip measurement (cm) (cm/cm)
- Units: Absolute ratio
- Covariates: sex

Heel bone mineral density

- Definition: Measured by Quantitative Ultrasound Index through the calcaneus
- Units: grams/centimeters^2 (g/cm2)
- Covariates: sex

Lung function

Forced expiratory volume in 1 second (FEV1)

- Units: Litres (l)
- Covariates: sex

FEV1/FVC ratio

- Definition: Ratio of Forced expiratory volume in 1 second (litres) to Forced Vital Capacity (litres)
- Units: Please state whether absolute ratio value or Z-scores are being provided.
- Covariates: sex

Cardiovascular

Systolic blood pressure (SBP)

- Definition: Automated reading
- Units: millimetres of mercury (mmHg)
- Adjustments: adjust for blood pressure medication. Applying a constant increase of 15mmHg in individuals taking medication.
- Covariates: sex

Blood measures

Low density lipoprotein (LDL) cholesterol

- Definition: blood measure
- Units: millimoles per liter (mmol/L)
- Adjustments: adjust for statin/cholesterol lowering medication by applying a relative reduction to the measured value of 40%.
- Covariates: sex

High density lipoprotein (HDL) cholesterol

- Definition: blood measure
- Units: millimoles per liter (mmol/L)
- Adjustments: cholesterol_med
- Covariates: sex

Triglycerides

- Definition: blood measure
- Units: millimoles per liter (mmol/L)
- Adjustments: cholesterol_med
- Covariates: sex

Glycated haemoglobin (HbA1c)

- Definition: blood measure
- Units: millimoles per liter (mmol/L)
- Adjusments: diabetes_status:insulin_medication
- Covariates: sex

Leptin

- Definition: blood measure
- Units: nanogram per milliliter (ng/ml)
- Covariates: sex

Insulin

- Definition: blood measure. Conversion factor 1 µIU/mL = 6.00 pmol/L.
- Units: pmol/L
- Adjustments: diabetes_status:insulin_medication
- Covariates: sex

Adiponectin

- Definition: blood measure
- Units: micrograms per milliliter (µg/ml)
- Covariates: sex

Calcium

- Definition: blood measure
- Units: millimoles per liter (mmol/L)
- Adjustments: cholesterol_med
- Covariates: sex

Vitamin D

- Definition: blood measure
- Units: nanomoles per liter (nmol/L)
- Covariates: sex

Immune markers

C-reactive protein (CRP)

- Definition: blood measure
- Units: milligrams per liter (mg/L)
- Adjustments: cholesterol_med
- Covariates: sex

Interleukin-6 (Il6)

- Definition: blood measure
- Units: picograms per millilitre (pg/ml)
- Covariates: sex

Sex hormones

Estradiol

- Definition: blood measure
- Units: picomoles per liter (pmol/L)
- Adjustments: apply constant for hormonal contraception use, and hormone replacement therapy
- Covariates: sex, age of menarche, parity (number of offspring) and age at menopause

Testosterone

- Definition: blood measure
- Units: nanomoles per liter (nmol/L)
- Adjustments: apply constant for hormonal contraception use, and hormone replacement therapy
- Covariates: sex, age of menarche, parity (number of offspring) and age at menopause

Sex hormone binding globulin

- Definition: blood measure
- Units: nanomoles per liter (nmol/L)
- Adjustments: apply constant for hormonal contraception use, and hormone replacement therapy

- Covariates: sex, age of menarche, parity (number of offspring) and age at menopause