Bristol demographics, replication plan

Tiffany Tu, Alejandro Ochoa

StatGen, Biostatistics & Bioinformatics — Duke University

2023-02-08 — NS U01 working group

Overview

- Bristol
  - All are cases
  - Small sample size (intersect with subtype, ancestry)
    - Only joint (all ancestries) analysis makes sense
  - Have to impute (most candidate loci are missing from raw)
- UK Biobank for controls
  - Overkill considering small number of cases
  - Too big to use it all (n=500,000) with same methods (GMMAT)
  - Can subsample, then do a joint analysis
  - If array genotypes, can impute too
  - WGS is more expensive, perhaps overkill
- GnomAD
  - Retrieve allele counts by ancestry, calculate joint p-value with LRT
  - If it can be automated, could test all suggestive loci this way!

Bristol demographics

Total $n = 590$ individuals without filters (age filter further reduces counts).

Sex	Count	%
Male	353	59.8
Female	237	40.2

Diagnosis	Count	%
SSNS	350	59.3
SRNS	172	29.2
NS unclassified	68	11.5

Race/Ethnicity	Count	%
White	402	68.1
Asian	84	14.2
Unknown	73	12.4
Black	16	2.7
Mixed	15	2.5

Bristol diagnosis subtypes

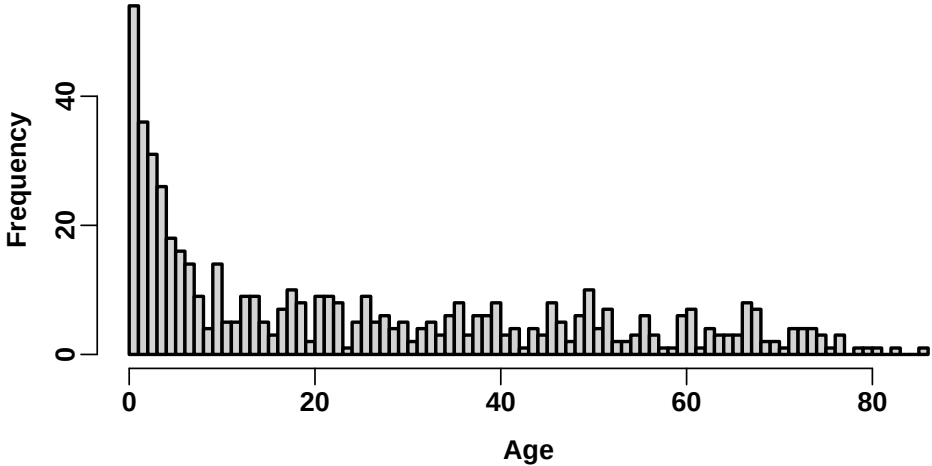
SSNS only

Race/Ethnicity	Count	%
White	229	65.4
Asian	57	16.3
Unknown	44	12.6
Black	7	2.0
Mixed	13	3.7

SRNS only

Race/Ethnicity	Count	%
White	122	70.9
Asian	23	13.4
Unknown	18	10.5
Black	8	4.7
Mixed	1	0.6

Age distribution



- 47 individuals missing age
- 294 individuals (49.8%) have age < 22

Bristol demographics, age < 22 only

Total $n = 294$ individuals

Sex	Count	%
Male	178	60.5
Female	116	39.5

Diagnosis	Count	%
SSNS	191	65.0
SRNS	96	32.7
NS unclassified	7	2.4

Race/Ethnicity	Count	%
White	174	58.8
Asian	56	19.0
Unknown	44	15.0
Black	11	3.7
Mixed	10	3.4

Bristol diagnosis subtypes, age < 22 only

SSNS only

Race/Ethnicity	Count	%
White	110	57.6
Asian	40	20.9
Unknown	25	13.1
Black	7	3.7
Mixed	9	4.7

SRNS only

Race/Ethnicity	Count	%
White	59	61.5
Asian	18	18.8
Unknown	14	14.6
Black	4	4.2
Mixed	1	1.0

UKBB costs (3,000 pounds = 3,620.18 USD)

Description	Tier 1	Tier 2	Tier 3
Core data • Questionnaires and physical measurements • Linked health data • Health Outcome phenotypes • Web-based questionnaires	✓	✓	✓
Assay data and enhanced measures • Biochemical and haematological assays • Measured and imputed genotypes • Other platform based assays • Other enhancements		✓	✓
Very large datasets • Imaging data * • Whole genome sequence data • Other large-scale assay data • Whole exome sequence data			✓ Via platform only
First 3 years - access to data with scheduled updates	£3,000	£6,000 (+£3,000 vs Tier 1)	£9,000 (+£3,000 vs Tier 2)
Additional Institution fee - each additional institution added to an application	£1,000 for first 3 years (£500 p.a. extension)		
Low & Middle Income Countries and Student Researchers ** - access to all datasets via the Research Analysis Platform (full fees apply to downloaded data)	£500 for first 3 years (£175 p.a. extension)		