PheWAS for Median Household Income in the HNP

Objective

Identify phenotypes strongly associated with income levels in the HNP.

Method

PheWAS using PheCodes for the HNP population with a DX_ID in the Renown EHR.

For each PheCode P_j, estimate the following via logistic regression:

$$P_{j,i} = \alpha + \gamma I_i + X_i'\beta + u_i,$$

where P is a binary indicator equal to one if patient i was observed to have a DX_ID with PheCode j in the EHR (0 otherwise), I_i is the log of median household income in the Census tract, and X_i is a vector of control variables.

Control variables include: binary indicator for female, age (in years), age-squared, dummy variables for each of the genetic ancestry categories (European is the excluded category), dummy variables for the patient being in Nevada and California (all other states are the excluded category), and dummy variables for each year 2019–2022 (where the year is equal to one if the patient joined the HNP in that year; 2018 is the excluded category).

Female status is determined genetically. For those that could not be assigned genetically, I use self-reported "BirthGender" from the consents. If still not assigned genetically or from the consent, I use self-reported sex in the EHR.

The p-value on the coefficient estimate for γ indicates the degree of statistical significance for household income in determining the log of the odds of observing phenotype j.

The PheWAS was limited to phenotypes that occurred 50 or more times in the population. This resulted in 1,278 phenotypes (out of the possible 1,866).

Data

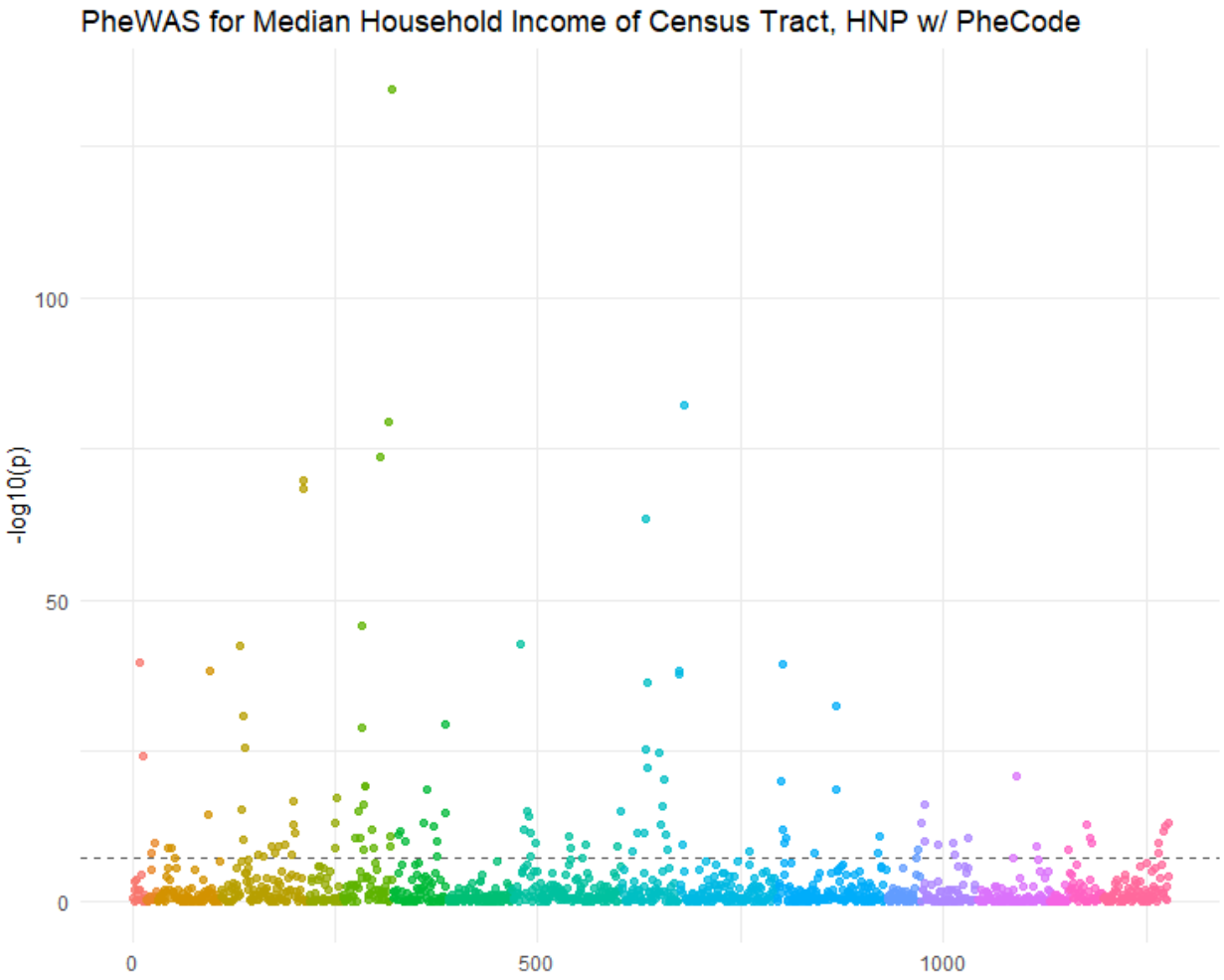
Sample of Renown recruits from the *monolith* (version Sept 20, 2022). Dropped one record where the same PAT_ID was associated with two PUIDs.

PheCodes taken from the *Gaiagnosis* database. Links PheCodes to DX_ID. Select for all HNP participants (by PAT_ID). The sample was determined as follows:

- N = 40,247 Renown recruits
- N = 37,304 with a PheCode.

The *Giagnosis* database includes about 4.3 million HNP patient-DX_ID records. About 8 percent of those were not matched to a PheCode. I excluded them for now.

Results



Top $-\log_{10}(P)$ phenotypes:

colref	log10p	phecode	phenotype
p533	134.3944461	318	Tobacco use disorder
p1021	82.28362098	525	Other diseases of the teeth and supporting structures
p528	79.41069635	316	Substance addiction and disorders
p515	73.60413926	306	Other mental disorder
p388	69.84680788	278.1	Obesity
p389	68.29901976	278.11	Morbid obesity
p953	63.51600256	496	Chronic airway obstruction
p488	45.8463538	296.22	Major depressive disorder
p761	42.70120006	401.1	Essential hypertension
p257	42.41224719	250.2	Type 2 diabetes
p57	39.6308626	1010	Other tests
p1185	39.45735076	585.31	Renal dialysis
p1013	38.31853707	522.5	Periapical abscess
p202	38.22407391	216.1	Screening for malignant neoplasms of the skin
p1008	37.71672549	521.1	Dental caries
p956	36.34675742	496.21	Obstructive chronic bronchitis
p1270	32.50194844	611	Abnormal findings on mammogram or breast exam
p261	30.81941872	250.24	Type 2 diabetes with neurological manifestations
p611	29.39037831	355.1	Chronic pain syndrome
p486	28.79888521	296.1	Bipolar
p263	25.59840084	250.3	Insulin pump user
p954	25.39086862	496.1	Emphysema
p974	24.83122293	509.1	Respiratory failure
p60	24.28299275	1013	Asphyxia and hypoxemia
p955	22.08783571	496.2	Chronic bronchitis
p1564	20.75662024	733	Other disorders of bone and cartilage
p985	20.29996583	512.7	Shortness of breath
p1182	19.90916848	585.1	Acute renal failure
p493	19.08833442	300.1	Anxiety disorder
p492	19.08285695	300	Anxiety disorders
p584	18.72563407	345.3	Convulsions
p1271	18.69644682	611.1	Abnormal mammogram
p449	17.10294242	288.2	Elevated white blood cell count
p373	16.70859481	276.41	Acidosis
p490	16.10355957	297.1	Suicidal ideation
p1404	16.01630235	681.5	Cellulitis and abscess of leg, except foot
p979	15.92231424	510	Other diseases of lung
p259	15.27129235	250.22	Type 2 diabetes with renal manifestations
p769	15.11587808	411.2	Myocardial infarction