

PEGS Phenotyping

Motsinger-Reif Group

1 Introduction

In the PEGS cohort we have chosen the disease phenotypes to examine based on a prevalence cutoff of about 10% of the study population. The phenotypes examined are Lower Gastrointestinal Polyps, Fibroids, Ovarian Cysts, Bone Loss, Migraines, Iron Deficient Anemia, Type 2 Diabetes, Asthma, Hypertension, High Cholesterol, cardiac arrhythmia, congestive heart failure, coronary artery disease, heart attack/myocardial infarction, stroke, and an outcome variable that combines multiple atherogenic-related outcomes, including angina, angioplasty, atherosclerosis, coronary artery disease, heart attack, and stroke. These disease phenotypes will be used when examining the impacts of genetics and the environment on individual's health. They were defined using literature sources and disease knowledge for inclusion and exclusion criteria. All questions asked to determine individuals phenotypes come from the PEGS Health and Exposure (H&E) Survey. In these definitions, we have not used the Exposome A, Exposome B or any of the other surveys in PEGS.

The diagnostic and exclusion criteria questions are all of the type "Has a doctor or other health care provider ever told you that you have" the disease of interest. While these are all self reported survey questions, the questions are phrased in a way that implies that they have been diagnosed professionally with the disease of interest. In our framework, only controls went through the exclusion process criteria in order to preserve case count from strict exclusion criteria. Cases are still kept no matter how they answered other questions unless otherwise noted.

2 Phenotypes

Lower GI Polyp Phenotype Creation: Original Lower GI Polyp Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Polyps in the colon or rectum." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1466 and original number of controls was 7828. Only controls went through the exclusion criteria process. Individuals were excluded from both the cases and controls if they had any cancer other than non melanoma skin cancer, had any of Lactose intolerance, Crohn's disease or Ulcerative Colitis^{1,2}. No Age cutoff was used. The final numbers of cases is 1466 and final number of controls is 6297.

Fibroid Phenotype Creation: Original Fibroid Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Fibroids, fibroid tumors, uterine fibroids or other benign uterine tumors." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1454 and original number of controls was 4771. Only controls went through the exclusion criteria process. Individuals were excluded from both the cases and controls if they had Ovarian Cysts, Endometriosis or uterine polyps. The final numbers of cases is 1454 and final number of controls is 3719.

Ovarian Cysts Phenotype Creation: Original Ovarian Cysts Cases were defined as those who answered yes to the question "Has a doctor or other Health Care provider ever told you that you have Ovarian Cysts or benign ovarian growth or neoplasm." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1218 and original number of controls was 5011. Only controls went through the exclusion criteria process. Individuals were excluded from both the cases and controls if they had Fibroids, Endometriosis or uterine polyps. The final numbers of cases is 1218 and final number of controls is 3725.

Bone Loss Phenotype Creation: Original Bone Cases were defined as those who answered yes to the question "Has a doctor or other Health Care provider ever told you that you have Bone loss, thinning in the bones, osteopenia or pre-osteoporosis" or if they had been diagnosed with Brittle Bones or Osteoporosis. Individuals who answered no to both questions were categorized as controls and those who answered yes to either question were considered cases. Originally, we had 1237 bone loss cases, 424 Osteoporosis cases, and 8034 controls, with overlap between the Bone loss and osteoporosis cases. If the individual was missing the question on either Bone loss or Osteoporosis they were excluded from the phenotyping. Controls were also excluded if they had been diagnosed with bone cancer. When bone loss and osteoporosis questions were combined and gone through the exclusion criteria, we were left with 1261 cases and 7940 controls.

Migraine Phenotype Creation: Original Migraine Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Migraine headaches (with or without aura)." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1643 and original number of controls was 7675. Only controls went through the exclusion criteria process. Individuals were excluded from both the cases and controls if they had brain cancer. No other neurological diseases were excluded. The final numbers of cases is 1643 and final number of controls is 7667.

Iron Deficient Anemia Phenotype Creation: Original Cases were defined as those who answered yes

to the question "Has a doctor or other health care provider ever told you that you have Iron Deficiency anemia." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1517 and original number of controls was 7804 .Only controls went through the exclusion criteria process. Individuals were excluded from both the cases and controls if they had pernicious anemia, sickle cell disease or leukemia. The final numbers of cases is 1517 and final number of controls is 7756.

Allergic Rhinitis Phenotype Creation: Original Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Allergic Rhinitis, Hay Fever or Seasonal Allergies." Individuals who answered no were categorized as controls and those who answered yes were considered cases. No other exclusion criteria was applied. The final numbers of cases is 3441 and final number of controls is 5037.

Type 2 Diabetes Phenotype Creation: Original Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Diabetes." Individuals who answered no were categorized as controls and those who answered yes were considered cases. The original number of cases was 1140 and controls was 7603. Controls who were diagnosed under 20 years old and those who were diagnosed with Gestational Diabetes were excluded. The age cutoff was used to exclude Type 1 diabetes cases. The final numbers of cases is 953 and final number of controls is 7603.

Asthma Phenotype Creation: Original Cases were defined as those who answered yes to the question "Has a doctor or other health care provider ever told you that you have Asthma" and "Has a doctor or other health care provider told you that you currently have Asthma". Individuals who answered no were categorized as controls and those who answered yes were considered cases. Cases were excluded if they had ever been diagnosed with COPD, idiopathic pulmonary fibrosis, or Tuberculosis. The final numbers of cases is 1228 and final number of controls is 7260.

Cardiovascular Disease Phenotype Creation: We considered eight self-reported cardiovascular outcomes from the health and exposure questionnaire high cholesterol, hypertension, cardiac arrhythmia, congestive heart failure, coronary artery disease, heart attack/myocardial infarction, stroke, and an outcome variable that combines multiple atherogenic-related outcomes, including angina, angioplasty, atherosclerosis, coronary artery disease, heart attack, and stroke. For the combined atherogenic-related outcome, we assigned participants to the case group if they answered 'YES' when asked if they have been diagnosed with any of the six atherogenic-related diseases or outcomes included in the combined variable. For individual outcome variables, we assigned participants to the case group if they answered 'YES' when asked if they have been diagnosed with a particular cardiovascular outcome. We assigned participants to the control group if they answered 'NO' to that specific cardiovascular-related traits. The final case counts were High Cholesterol: 3084, Hypertension: 3456, Cardiac Arrhythmia: 559, Congestive Heart Failure: 183, Coronary Artery: 311, Heart Attack: 289, Stroke: 227, and Combined Atherogenic: 818.