|  |  |
| --- | --- |
| **eMERGE Network: Proposal for Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT234 |
| **Submission Date** | 5/22/2017 |
| **Project Title** | GWAS of Medicare Risk Adjustment Model scores |
| **Tentative Lead Investigator** *(first author)* | Scott Hebbring |
| **Tentative Senior Author**  *(last author)* | Josh Denny |
| **All Other Authors** | eMERGE collaborators |
| **Sites Involved** | Marshfield Clinic  Vanderbilt  Other eMERGE sites |
| **Background / Significance** | There is great public interest in understanding the relationship between human genomics and healthcare utilization. As a result, state and federal agencies have enacted legislation to prevent genetic discrimination (e.g., GINA). Federal and private payers apply predictive models to evaluate a population’s future risk for healthcare utilization to estimate future costs and set premiums. There are multiple published predictive models used throughout the insurance industry but one of the most standardized models include Medicare Risk Adjustment Model (MRAM) that can be applied to individuals over 65 years of age. A MRAM score can be calculated in an individual by using a patient’s ICD9 coding history, age, and sex.  C:\Users\hebbrins\AppData\Local\Avecto\DragAndDrop\GWAS_combined_EX_ALL_HQ_health_economic_tran_LRT_Manhanttan_gp_cf0.1_thin0.8.tiff  Scott Hebbring recently conducted a pilot GWAS on MRAM scores in 5,500 PMRP participants with ICD9 coding data up to age 65 years (see Figure). In short, no association passed genome-wide significance and no suggestive association (p<1E-5) included known disease causing variants. This may not be a surprising results given most common variants associated with human disease have weak effects. Regardless, this null result may have an important impact on how genomic data is used to evaluate healthcare utilization. To follow-up these findings, we propose to expand the study across multiple eMERGE sites and to further evaluate rare pathogenic variants on MRAM scores. |
| **Outline of Project** | GWAS on MRAM   1. GWAS of common variants 2. Focus on common disease causing SNPs (i.e., GWAS Catalogue SNPs) 3. Focus on known and rare pathogenic variants genotyped directly by Illumina Exome array or equivalent data sources (exome sequencing data)   Generalized linear mixed model (GLMM) and/or burden test   1. GLMM to quantify how much common variants contributes to MRAM variance 2. GLMM to quantify how much common disease causing SNPs (GWAS Catalogue SNPs) contributes to MRAM variance 3. Burden test to evaluate if increase number of disease causing (GWAS Catalogue SNP and/or rare pathogenic variants) alleles are associated with MRAM |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | Phenotype/demographic data   1. \*ICD9 coding data for those with available data beyond, but must include, 60-65 years of age. Variables needed include ICD9 code and age at each ICD9 code. Demographic data includes sex and race.   Genomic data   1. \*Imputed and QCed SNP array data in patients that meet phenotypic requirements (pulled from eMERGE I and II or equivalent). 2. QCed Exome SNP array or sequencing data (eMERGE III) in patients that meet phenotypic requirements. |
| **Desired Data** | Genotype, phenotype, and demographic data |
| **Planned Statistical Analyses** | 1. MRAM will be applied to ICD9 coding data up to age 65, other age cutoffs may be considered 2. Box-Cox transformation of MRAM 3. GWAS by linear regression, covariates include length of EHR data, eMERGE site, and principal components. 4. GLMM: evaluate genome and GWAS Catalogue SNPs 5. Burden analysis: evaluate GWAS Catalogue SNPs 6. Identify all disease risk SNPs from GWAS Catalogue 7. Estimate allele dosage for disease causing variants. When multiple SNPs are part of a LD block (r^2>0.2), a permutation routine will be implemented to randomly select SNPs used for estimating allele dosages. Additional considerations may be needed when multiple diseases map the same SNP/LD block. |
| **Ethical Considerations** | There are no inherent ethical issues with the proposed study since all individuals will be de-identified. Conversely, MRAM is used by healthcare payers to evaluate future healthcare costs for the population they serve and to set premiums. Given the topic being studied, positive or negative results could lead to inappropriate and biased conclusions by others. For example, a positive or negative result could be used politically to justify weakening current laws that prevent genetic discrimination (e.g., GINA). |
| **Target Journal** | High impact |
| **Milestones\*\*** | Completion within one year  Month 1: Obtain genetic and phenotypic data across participating eMERGE sites  Month 3: Calculate MRAM across eMERGE sites  Month 5: Initial GWAS of common variants including GWAS Catalogue SNPs  Month 8: Evaluations of rare variants, burden test, and GLMM  Month 10: First draft of manuscript  Month 11: Second draft of manuscript |

***\*\**** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*