|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT449 | |
| **Submission Date** | April 25, 2022 | |
| **Project Title** | A Polygenic Risk Score (PRS) for Prostate Cancer: retrospective validation via EHR algorithm & prospective implementation for clinical risk assessment | |
| **Tentative Lead Investigator** *(first author)* | Akshar Patel | |
| **Tentative Lead Investigator Email Address** | adam.gordon@northwestern.edu, akshar.patel@northwestern.edu | |
| **Tentative Senior Author**  *(last author)* | Adam Gordon | |
| **All Other Authors** | Jennifer Pacheco, Luke Rasmussen, Laura Rasmussen-Torvik, Philip Silberman | |
| **Sites Participating** | NU, any interested eIV sites | |
| **Background / Significance** | Prostate Cancer is one of ten conditions for which a Polygenic Risk Score (PRS) will be prospectively returned to participants in primary care as part of the fourth phase of eMERGE. Here we describe the selection, adaptation, and validation of the PRS, as well as the development of a novel EHR algorithm to identify Prostate Cancer cases within biobank data. We will also describe the deployment of the score as a tool for clinical risk assessment in a primary care setting. | |
| **Outline of Project** | * PRS selection/adaptation   + Lit/catalog review of potential scores, comparison of published performance and approaches to ancestry * Case/control validation in eI-III data   + ICD-based EHR algorithm for case/control/exclude (and validation via chart review)   + Calculation of PRS in eI-III GWAS data   + Association testing, selection of risk threshold, standardized PRS metrics     - Sub-analysis: age of onset, disease severity * Prospective deployment   + Adaptation to clinical pipeline   + Evaluation of ancestry correction methods   + Prostate Cancer-specific GIRA content | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Geocoding 2015 ACS variables  Other: Case/Control status |
| **Other Desired Data *(Available from participating sites)*** | * Clinical data linked to disease severity (e.g. PSA, Gleason score, clinical grade/stage) in eI-III cases, if available * We are very interested in applying our standardized\* EHR algorithm + PRS to biobank data outside of eI-III for additional validation, if available. (\*We have OMOP CDM queries for our algorithm that we can share, & we’re happy to help with any tweaks needed to make that OMOP SQL code run at your site) | |
| **Desired Genetic Data** | eMERGE I-III Merged set (HRC imputed, GWAS)  eMERGE PGx/PGRNseq data set  eMERGEseq data set (Phase III)  eMERGE Whole Genome sequencing data set  eMERGE Exome chip data set  eMERGE Whole Exome sequencing data set  Other (not listed above): | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | * Calculation of PRS and standardized metrics in eI-III data (aligned with larger PRS workgroup paper) * Association testing between PRS and markers of disease severity, if available | |
| **Ethical Considerations** | None | |
| **Target Journal** | AJHG, PLoS Genetics, JNCI | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | May 2022 MCS approved  May 2022 Outline to authors  July 2022 Draft to Authors  August 2022 Submission to journal | |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants