|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT438 | |
| **Submission Date** | November 10, 2021 | |
| **Project Title** | A Novel Metadata Framework for Phenotyping Algorithms | |
| **Tentative Lead Investigator** *(first author)* | Matthew Spotnitz | |
| **Tentative Lead Investigator Email Address** | mes2165@cumc.columbia.edu | |
| **Tentative Senior Author**  *(last author)* | Chunhua Weng, Karthik Natarajan, Wei-Qi Wei | |
| **All Other Authors** |  | |
| **Sites Participating** | All eMERGE sites | |
| **Background / Significance** | Electronic Phenotypes provide inadequate descriptions to help users identify the algorithms that will retrieve patient cohorts that match their expectations. For example, there are more than a dozen Type 2 diabetes mellitus phenotyping algorithms that retrieve partially overlapping patient cohorts. | |
| **Outline of Project** | 1. Define a metadata framework to enrich cohort descriptions (e.g., provide more information than just “type 2 diabetes mellitus”) 2. Evaluate the metadata framework based on domain expert input 3. Apply the metadata framework to annotate all electronic phenotypes 4. Iterate the steps 1-3 as needed 5. Evaluate if the metadata improves accuracy and efficiency for retrieval of cohorts | |
| **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)*** | *Phenotypes:*  *EHR derived phenotypes and the associated data elements (diagnosis codes, procedure codes, medication codes, age, sex, race, ethnicity).* | |
| **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 All phenotypes  No | |
| **Planned Statistical Analyses** | None | |
| **Ethical Considerations** | None at this time | |
| **Target Journal** | JAMIA, Journal of Biomedical Informatics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 12/2021: Develop phenotype framework and begin annotation  02/2022: Finish phenotype annotation  03/2022: Submit Manuscript | |

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

* Demographics: sex, age, race, ethnicity
* Codes: ICD-09/10, CPT, LOINC, ATC, SNOMED-CT, OMOP CDM, Phecodes, or other codes used for phenotyping