|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT327 | |
| **Submission Date** | 02/07/2019 | |
| **Project Title** | A Study of Phenotype Algorithm Portability | |
| **Tentative Lead Investigator** *(first author)* | Ning “Sunny” Shang, Cong Liu | |
| **Tentative Senior Author**  *(last author)* | Chunhua Weng, George Hripcsak | |
| **All Other Authors** | Jennifer Allen Pacheco, Luke Rasmussen, and all other sites participating in the effort | |
| **Sites Participating** | We propose a network-wide study (all sites invited to participate). The analyses will be led by Columbia University. So far the participating sites include Columbia University and Northwestern University. | |
| **Background / Significance** | eMERGE studies have demonstrated that knowledge-engineering based phenotyping algorithms can be transferrable to different institutions, which have heterogeneous data models, EHR systems, and local practice patterns [1]. To ease the efforts of implementation within and outside of the eMERGE network, developing portable phenotyping algorithms is desired [2,3]. To assist portability, information model has been explored [4,5]. However, how the portability varies among different phenotypes has not been assessed yet. We aim to quantify the portability of eMERGE algorithms by calculating how many communications and customizations are needed to implement the algorithm. In this study, we will measure the portability of 55 eMERGE algorithms, and subsequently will assess how OMOP CDM can help alleviate the portability burden. | |
| **Outline of Project** | 1. Two individuals with expertise in implementation of algorithms will review the algorithms and summarize the customization tasks for implementing eMERGE phenotypes. 2. Survey the participating eMERGE sites to review and confirm the proposed customization tasks and to assess how much effort could be reduced by using the OMOP solution. 3. Data analysis. 4. Prepare and submit the manuscript. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI  **NONE** | Common Variable Labs  Common Variable Meds  Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*   * Final versions of 55 eMERGE algorithms from Phekb. * Results from the survey. | |
| **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):  **NONE** | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list: Cdiff, CaMRSA, Carotid Artery Atherosclerosis Disease (CAAD), Benign Prostatetic Hyperplasia, Appendicitis, Colon Polyp, Extreme Obesity, MACE Statin, Asthma, Age-related Macular Degeneration, CKD, Cardiorespiratory Fitness, ADHD, Atopic Dermatitis, GERD, Glaucoma, Ocular Hypertension, Heart Failure, Diverticulosis, Venous Thromboembolism, Autism, Early childhood obesity, Abdominal Aortic Aneurysm (AAA), ACE inhibitor/cough, Zoster, T2DM, Cataract, Dementia, Peripheral Arterial Disease (PAD), QRS, RBC, WBC, ResHTN, Height, Hypothyroidism, Lipids, Diabetic Retinopathy, Adult familial hypercholesterolemia (FH), Colorectal Cancer (CRC), Migraine (Peds and Adult), Epilepsy, Chronic Rhinosinusitis, Chronic Kidney Disease (eGFR, proteinuria), Contrast Induced Nephropathy, Hearing Loss, Rheumatoid arthritis, Fatty Liver condition= (NAFLD/NASH-Alcoholic), Intellectual disability, Arrhythmias EKG Intervals, Ovarian/Uterine Cancer, Autoimmunity, Metformin response (PGx), Periperal arterial disease, Pneumonia, Anxiety  No | |
| **Planned Statistical Analyses** | Descriptive summary of the number of tasks and the agreement among participants. | |
| **Ethical Considerations** | NONE | |
| **Target Journal** | JBI | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Total Duration of the study: 2 months  Completion of study design/approvals: Feb 2019  Summarize localization tasks for portable phenotyping implementation: Feb 2019  Survey results: Feb 2019  Draft of manuscript to authors: Feb-March 2019  (with a possible earlier design paper)  First submission: April 2019 | |

**\*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

Reference

1. Denny JC, Crawford DC, Ritchie MD, et al. Variants Near FOXE1 Are Associated with Hypothyroidism and Other Thyroid Conditions: Using Electronic Medical Records for Genome- and Phenome-wide Studies. The American Journal of Human Genetics 2011;89:529–42. doi:10.1016/j.ajhg.2011.09.008
2. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013;15:761–71. doi:10.1038/gim.2013.72
3. Zhong Y, Rasmussen L, Deng Y, et al. Characterizing Design Patterns of EHR-Driven Phenotype Extraction Algorithms. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE 2018. 1143–1146.
4. Mo H, Pacheco JA, Rasmussen LV, et al. A Prototype for Executable and Portable Electronic Clinical Quality Measures Using the KNIME Analytics Platform. AMIA Jt Summits Transl Sci Proc 2015;2015:127–31.
5. Pacheco JA, Rasmussen LV, Kiefer RC, et al. A case study evaluating the portability of an executable computable phenotype algorithm across multiple institutions and electronic health record environments. J Am Med Inform Assoc 2018;25:1540–6. doi:[10.1093/jamia/ocy101](https://doi.org/10.1093/jamia/ocy101)