|  |  |
| --- | --- |
| **eMERGE Network: Proposal for Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT329 |
| **Submission Date** | Feb 14, 2018 |
| **Project Title** | Framework for Assessing EHR-Based Phenotype Algorithm Complexity |
| **Tentative Lead Investigator** *(first author)* | Yuan Luo |
| **Tentative Senior Author**  *(last author)* | Wei-Qi Wei |
| **All Other Authors** | Jennifer Pacheco, Luke Rasmussen, any other eMERGE NLP collaborators explicitly expressing interest |
| **Sites Involved** | Northwestern, any other interested institutions |
| **Background / Significance** | As more NLP-based phenotypes have been proposed, developed and implemented across multiple institutions as part of eMERGE (and beyond), we have recognized that not all NLP phenotype algorithms are created equal. The time to develop a single NLP phenotype algorithm within an institution, as well as the time for another institution to implement the same NLP phenotype algorithm may vary greatly.  While we understand that some NLP phenotype algorithms are more “complex” than others, this notion of complexity can be measured in different ways. First, a certain amount of work needs to go into pulling the clinical notes that are fed into the NLP phenotype algorithm. Second, there is a notion of complexity in the implementation of the NLP phenotype logic using clinical notes. Finally, the external resources requested represent another level of complexity.  In this work, we propose to develop a framework by which a complexity metric of an NLP phenotype algorithm may be measured, and provide context around the interpretation of this metric. This builds upon existing work with clinical quality measures [1] and portable phenotype algorithms [2], as well as an earlier analysis of eMERGE phenotypes [3,4]. We will assess the metric across implementations of existing eMERGE 1-3 NLP phenotypes, which will also allow us to better understand any changes in complexity of phenotypes over time.  [1] Dorr DA, et al. “From simply inaccurate to complex and inaccurate: complexity in standards-based quality measures”  [2] H Sharma, C Mao, et al. “Developing a Portable Natural Language Processing Based Phenotyping System”  [3] Conway M, et al. “Analyzing the Heterogeneity and Complexity of Electronic Health Record Oriented Phenotyping Algorithms”  [4] Thompson WK, et al. “An Evaluation of the NQF Quality Data Model for Representing Electronic Health Record Driven Phenotyping Algorithms” |
| **Outline of Project** | 1. Further refine the proposed complexity metric (initial version presented in Desired Variables) 2. Using the complexity metric definition, use a structured form to quantify/classify the associated attributes for a phenotype definition (number of phenotypes depends on how many individuals are involved and interested – ideally a minimum of 20) 3. Analyze results of the classification |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | \*For each NLP phenotype algorithm analyzed for this study, we will classify:   * Input complexity   + For each data element that is required to calculate a phenotype, classify:     - Type of clinical notes (e.g., pathology reports, discharge summaries)     - Value set used to identify data element       * Vocabulary(ies) used       * Number of total elements after full expansion (if not fully expanded)     - Anticipated modality of the data (free text, handwritten notes requiring OCR)     - Is associated provider information needed?     - Are structured EHR data values (e.g., laboratory value, medication dose/route) needed? * Algorithm complexity   + Estimated cyclomatic complexity of pseudocode   + Max depth of Boolean operators   + Number of temporal relationships   + Number of assertion/exclusion operators   + Is temporal resolution needed?   + Is named entity/concept recognition needed?   + Is assertion classification needed?   + Is word sense disambiguation needed?   + Is semantic relation or event frame extraction needed?   + Is regular expression patterns needed?   + Is mixture of regular expression, concept, and relation patterns needed?   + Are different NLP pipelines used?   + Is algorithm recursively dependent on other NLP phenotype algorithms? * External resource complexity   + Is general knowledge sources (e.g., wordnet) needed?   + Is domain specific knowledge sources (e.g., UMLS) needed?   + Is customized/home constructed knowledge sources needed?   \*Each co-author involved is asked to classify at least two phenotype algorithms to ensure we have sufficient coverage. |
| **Desired Data** | * NLP Phenotype algorithm definition as pseudocode or other higher-level representation (e.g., HQMF) * Data dictionary(ies) |
| **Planned Statistical Analyses** | Summary statistics showing variation across the axes and domains of complexity that we measure. |
| **Ethical Considerations** |  |
| **Target Journal** | JAMIA |
| **Milestones\*\*** | Feb 2019 – Submit concept sheet  End of May 2019 – Complete review and proposal of complexity metric  End of July 2019 – Complete review of phenotypes using metric (co-authors)  End of September 2019 – Complete analyses of results  End of November 2020 – First draft of article complete  End of January 2020 – Review and feedback sent by co-authors  End of February 2020 – Second draft completed  End of March 2020 – Submit manuscript |

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