Individuals and Organizations Completing Research in the lab:

Chunhua Weng
Associate Professor
Department of Biomedical Informatics
Columbia University

Scope of Work:

Background:

Clinical trials are the gold standard for generating new high-quality medical evidence. Many clinical trials are designed to emphasize internal validity, and some decisions may compromise external validity. When a clinical trial has limited generalizability, the study results can be difficult to translate to the real-world population that would otherwise be the users of the study information. Little is known about how the populations of related trials collectively represent the real-world population with the condition despite the fact that such knowledge is important for patient-centered outcomes research (PCOR).

Research Objectives and Aims:

1. To develop methods for profiling patient populations from longitudinal observational data
2. To inform clinical trial patient selection with population summaries

Scope/Proposed Approach:

We have previously developed and published a distribution-based method[1] for comparing electronic health record-based patient population in one hospital (New York Presbyterian) to clinical trial patient populations, as defined by inclusion and exclusion criteria available from the Clinicaltrials.gov.

We developed the Generalizability Index for Study Traits (GIST) as a metric to evaluate the relative generalizability of a study characteristic shared by a set of clinical trials to a real-world population. GIST is the sum across all intervals of the proportion of trials including patients in that interval, multiplied by the proportion of patients in the real-world population observed in that interval.

$$GIST = \sum_{i=1}^{N} \frac{\sum_{j=1}^{T} I(i_{low} <= w_j <= i_{high})}{T} * \frac{\sum_{k=1}^{P} I(i_{low} <= y_k <= i_{high})}{P}$$

Where N is the number of distinct intervals of the study trait, T is the number of trials, P is the number of patients in the population, w_j is the inclusion interval for the j^{th} study, such that an indicator I can be defined when the j^{th} study interval subsumes the i^{th} interval low and high boundary threshold, and y_k is the observed value of the

characteristic for the k^{th} patient such that an indicator I can be defined when the k^{th} patient's value falls within the i^{th} interval. The GIST metric is on a 0 to 1 scale that characterizes the proportion of real-world population that would be eligible across the clinical trial studies, with 1 being perfectly generalizable (all patients would be eligible for all studies), and 0 being completely not generalizable (no real-world patients would be eligible for any studies).

This published method has two limitations: (1) use of one institution's patient-level observational data; and (2) use of a single phenotypic trait. In this study, we plan to assess the generalizability of this method of comparing clinical trial inclusion criteria to heterogeneous observational data sets and to extend this method by using multiple phenotypic traits simultaneously.

Of note, this work is not directly intended to compare the results of RCTs with the results of observational studies, but simply to provide perspective of the extent to which the RCT populations (as inferred from the study inclusion criteria in clinicaltrials.gov) may differ from the real-world population as one potential cause of difference that potentially explains study result differences. This work does not evaluate the true clinical trial population composition, but only the potential composition as defined by the inclusion criteria, as the true population information is not currently available in a systematic, structured manner, and we have developed NLP processes to extract the inclusion criteria into a structured form. Development of a multivariate generalizability metric will expand the current work to determine what proportion of patients in the real world would satisfy multiple inclusion criteria from the clinical trials.

Impact:

1. Methods for profiling and summarizing patient populations from longitudinal observational data
2. Methods for comparing real-world patient populations with study criteria of multiple related clinical trials on the same medical condition

Experience:

Chunhua Weng, Associate Professor, Department of Biomedical Informatics, Columbia University

Timeline: *24 months (August 15, 2014- August 14, 2016)*

References:

1. Weng C, Li Y, Ryan P, Zhang Y, Gao J, Liu F, Bigger JT, Hripcsak G, A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records, Applied Clinical Informatics, Vol. 5: Issue 2 2014, 463-479.