Category: Trial Recruitment and Community Outreach and Engagement – Work in Progress

Machine Evaluation of Catchment Area Relevance Through Text Mining

P. Arlen, J. Chakko, B. Mahal, G. DeGennaro

Sylvester Comprehensive Cancer Center, University of Miami Health System

1. Background

The University of Miami Sylvester Comprehensive Cancer Center (Sylvester) is located in South Florida, with a catchment area that represents the most racially, ethnically, and geographically diverse region in the U.S. Unfortunately, the area's tumor burden is also significant and with many notable disparities, necessitating a prioritization of trials within Sylvester's catchment area. These trials address the needs of the population Sylvester serves by targeting cancers that are locally prevalent, such as prostate and breast; comprise a special population, such as firefighters; are of local concern to those who live in South Florida, such as environmental exposures; or are subject to disparities in treating diverse populations, such as infection with human papillomavirus (HPV). Focusing on these needs of our catchment area is vital to effectively serving our patients.

2. Goals

Our goal is to assess the catchment area relevance of Sylvester clinical trials with a rubric that measures multiple parameters. We plan to automate this process and supply results to investigators, site disease groups (SDGs), and study team members.

3. Solutions and Methods

The rubric ranges in score from 0-8, with higher values indicating greater relevance of trials to catchment area criteria. First, a knowledgeable person assigns a catchment area score to a sample of trials. These scores are used by the machine to evaluate its own performance. The machine searches for key phrases related to each rubric criterion as shown below.

4. Outcomes

Developing an algorithm that assigns catchment score creates a new data point for considering and prioritizing trials. The rubric and automated scoring algorithm perform best on objective and easily accessible signifiers, such as a trial's disease indication. Determining if a trial addresses disparities is the most challenging criterion to score. When evaluating a subjective criterion, human and machine scorers alike benefit from additional scored examples (for example, training) and clearly defined rules.

The scoring algorithm was applied to more than 300 oncology trials available at the University of Miami. Preliminary analysis of the results showed the algorithm correctly flagged trials studying a prevalent cancer in 92 percent of cases (type II error = 8 percent) and correctly flagged trials that do not in 91 percent of cases (type I error = 9 percent). This shows that more training examples are required to capture relevant trials that the algorithm currently mislabels.

5. Lessons Learned and Future Directions

Cancer centers wish to serve the needs of their patient populations by opening trials that are relevant to their catchment areas. Investigators can select or design relevant trials more easily when provided rapid access to appropriate metrics. Text mining can be applied to the eligibility criteria of trials to extract new catchment area score data points. Creating a robust key phrase bank is vital to ensuring the scoring algorithm is objective, fair, and accurate. The scoring process must be clear and understood before any attempts at automation are made.

Category: Trial Recruitment and Community Outreach and Engagement – Work in Progress

Figure: