Machine Evaluation of Catchment Area Relevance

A Text Mining Application

Joseph Chakko, MSc; Brandon Mahal, MD; Geoffrey DeGennaro, MSc; Philip A. Arlen, PhD
Sylvester Comprehensive Cancer Center at the University of Miami Miller School of Medicine

SYLVESTER COMPREHENSIVE CANCER CENTER
UNIVERSITY OF MIAMI HEALTH SYSTEM

UNIVERSITY OF MIAMI
MILLER SCHOOL of MEDICINE

Background and Purpose

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 should address the needs of the population Sylvester serves by targeting cancers that are locally prevalent, such as prostate and breast; comprise a special population, including 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 papilloma virus (HPV). Focusing on these needs of our catchment area is vital to serving our patients effectively.

The purpose of this project is to create a new identifying feature in our trial database. After storing the evaluations, researchers will be able to filter and run metrics on trials that have been labeled as Catchment Area Relevant.

Goals

Create a rubric to standardize the Catchment Area Relevance evaluation process

Evaluating catchment area relevance should be a clearly defined process. Any means of evaluating a study should produce the same score each time. To this end, our goal is to create a rubric that measures multiple parameters and that can be applied by knowledgeable users without additional training.

Automate trial evaluation by implementing a scoring algorithm

The evaluation process will be automatically handled by computers. The logic of the rubric can be coded in programming languages like Java and SQL. The program's execution mimics that of a human evaluator, making it easy to understand and adapt.

Communicate results throughout the organization

The results will be reportable to investigators, study team members, Site Disease Groups (SDGs), protocol review committees, and leadership.

Methodology

- First, a knowledgeable person uses the rubric to assign a catchment area score to each trial in a sample selection. The results of this scoring process are used as a guide when writing and adjusting the program code.
- The machine implementation parses protocol titles and abstracts based on key phrases and awards relevance points. Creating and maintaining a key phrase bank for each rubric criterion is the main driver of the machine evaluator's success.
- The machine parses a trial's title using each criterion's key phrase bank. If the machine finds a match, the trial is awarded points for meeting that criterion.

Catchment Area Relevance Rubric

Criterion	Key Phrase Examples	Points Awarded
Prevalent Cancer	Prostate, Melanoma, Breast	1
Special Population	Firefighters, Haitians, Hispanics	2
Regional Impact	Exposure, Smoking, Microbiome	2
Address Disparity	Community, Outreach, Equity	3

Scoring Examples

Oral Tongue Squamous Carcinoma – Retrospective Study on Gender, Age and Ethnic Disparities
[Addressing Disparity: 3 Points]

Multimodal treatment of Advanced Prostate Cancer using combined local and Systemic Therapy
[Prevalent Cancer: 1 Point]

Examining the Association of Polybrominated Diphenyl Ethers (PBDE) and Thyroid function of South Florida Firefighters
[South Florida: 2 Points; Special Population: 2 Points]

Results

The scoring algorithm was applied to more than 300 oncology trials available at Sylvester. In order to determine accuracy, the output was compared against the score provided by a human evaluator. This comparison of results showed the algorithm correctly flagged trials studying a prevalent cancer in 92% of cases (type II error = 8%) and correctly flagged trials that do not in 91% of cases (type I error = 9%).

Both the rubric and scoring algorithm are most effective when evaluating objective and easily accessible identifiers, such as diagnosis. On the other hand, a subjective criterion, such as whether a trial addresses disparities in the community, is more difficult for all evaluators. In these cases, human and machine scorers alike benefit from additional scored examples (for example, training) and clearly defined rules.

Implementing an algorithm that assigns a score for catchment area relevance creates a new data point for analysis. Database queries can reference the score to filter trials based on catchment area relevance, increasing visibility and establishing a quantifiable metric for strategic planning.

Lessons Learned

The automated scoring algorithm is straightforward enough to adapt to a variety of circumstances. Before any attempts at automation are made, the scoring process must be clearly understood. After all, the scoring algorithm can only be as good as the rubric that serves as its guide.

The algorithm has key requirements that play a large role in achieving accurate results. First is the key phrase bank for each criterion. The phrase banks must be comprehensive and kept current to ensure good performance that is objective, fair, and accurate.

Second is hand-scored examples by experienced evaluators. The algorithm's output is compared to the hand-scored examples to determine the accuracy of the program. These examples are time-consuming to create and should be prioritized.