A Practical Method for Transforming Free-Text Eligibility Criteria into Computable Criteria

Samson Tu, MS[1], Mor Peleg, PhD[1,2], Simona Carini, MS[3], Michael Bobak, MS[3], Jessica Ross MD[3], Daniel Rubin, MS, MD,[1] Ida Sim, MD, PhD[3]

[1]Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA

[2]Department of Management Information Systems, University of Haifa, Haifa, Israel

[3]University of California, San Francisco, CA

Abstract

Background: Human studies are the most important source of evidence for advancing our understanding of health and disease, and forms the basis for improving health care. It is therefore crucial to optimize the design, execution, and use of human studies. Making eligibility criteria, which define the study population, computer-interpretable could facilitate such optimization. At the design stage, study investigators could query a library of computable criteria to help define their study population by comparing the content and selectivity of their criteria to those of related studies. At the execution stage, investigators could query electronic health records to find potentially eligible subjects. Finally, at the usage stage, providers at the point of care could query for studies that enrolled patients similar to theirs. Formalizing eligibility criteria in a computer-interpretable language, however, is an extremely labor intensive task.

Objective: We aimed to incrementally capture the semantics of eligibility criteria from free-text into a standardized, computable format.

Methods: We previously defined the Eligibility Rule Grammar and Ontology (ERGO). For this work, we defined an intermediate representation called ERGO Annotation that is informed by both the complexity of natural language and the requirements for computability. ERGO Annotation models three statement types: 1) simple statements making single assertions, 2) statements about quantitative comparisons, and 3) complex statements, which are simple and/or comparison statements joined by Boolean connectives or semantic connectors (e.g., evidenced_by). We then combined various natural language processing tools (LexAccess, Open Biomedical Annotator, Open NLP Parser) with custom heuristics to partially automate the encoding of free-text eligibility criteria into ERGO Annotation.

Results: We selected 4 trials from ClinicalTrials.gov and compiled a test set of 60 eligibility criteria comprising 110 simple and 14 comparison criteria. We performed three evaluations. First, our semi-automated process achieved approximately 70% full or partial match to hand-coed versions of the test criteria in ERGO Annotation. Second, we illustrate the use of ERGO Annotation in Protégé OWL and with a relational database to: 1) structure and make queryable a library of eligibility criteria, 2) screen patients for study eligibility, and 3) search for studies with particular study populations. Finally, we analyzed 1000 randomly selected eligibility criteria from ClinicalTrials.gov to estimate the proportion of common criteria that can be captured in ERGO Annotation.

Conclusions: We demonstrate a practical method for incrementally transforming free-text eligibility criteria into ERGO Annotation to derive value for optimizing the design, execution, and use of human studies.