Standardizing Racial and Ethnic Categories to Assess Equity of Clinical Guideline Evidence

Jay D. S. Franklin¹, Miao Qi¹, Shruthi Chari¹, Morgan A. Foreman², Kristin P. Bennett, PhD¹, Oshani Seneviratne, PhD¹, Amar K. Das, MD, PhD², Deborah L. McGuinness, PhD¹
¹Rensselaer Polytechnic Institute, Troy, NY; ²IBM Research, Cambridge, MA

NOTE: This document will not be submitted directly; AIDR requests all submissions be entered in text form through the EasyChair web interface. I’ll copy and paste the text into the online text box when I submit, I’m just using Overleaf for convenience.

1 Abstract

Clinical practice guidelines are widely used to support evidence-based healthcare decisions. However, these guideline recommendations may not be equitable since the evidence from cited clinical studies may be racially and ethnically underrepresented. We investigate the equitability of clinical guidelines by analyzing whether the combined cohort population from referenced studies, which we call the ‘metacohort,’ is representative of the population that has the medical condition. We build on methods we have developed for study cohort modeling, knowledge extraction, and equity computation to allow for measure of equity of race and ethnicity in a guideline metacohort against those of a benchmark population derived from the nationally representative National Health and Nutrition Examination Survey (NHANES).

We selected 13 US-based clinical trials cited in the cardiovascular comorbidity chapter of the American Diabetes Association (ADA) Standards of Medical Care guideline 2019 as our sample set. We then compute the racial and ethnic equity between the resulting metacohort and the NHANES population. To extract study cohort information reported in tables in research publications, we have previously developed a study cohort extraction pipeline. The pipeline begins by using the IBM Corpus Conversion Service to extract tabular data from publication PDFs. It then identifies hierarchies of table rows, annotates text segments with medical terms or other knowledge elements, and semantically relates knowledge elements to one another. The final structure the pipeline outputs is an RDF knowledge graph modeling the study cohort information and conforming to the Study Cohort Ontology, ideal for representing the oftentimes complex information within a study cohort table.

The extracted study cohort information cannot yet be combined into one metacohort and compared to a representative population, as different publications may report race and ethnicity categories that are incompatible with one another or with the representative NHANES population. Hence, all such information must first be aligned and standardized where possible, at which subpopulations can be confined and direct comparisons can be made. We have begun work on an extension module to the Human Health Exposure Analysis Resource (HHEAR) ontology. We are using this extension module to create a unifying standard between the RDF knowledge graphs created by the pipeline and the benchmark population calculated from NHANES. We are implementing an algorithm to account for the differences between HHEAR and NHANES and create metacohorts that can be used for comparison.

Once standardized, the race and ethnicity categories of the metacohort can be analyzed by a previously implemented trial equity algorithm that can quantify the racial and ethnic equity of the metacohort against a target representative population. Hence, the equity of recommendations in the clinical practice guideline for specific underserved populations can be evaluated. Our existing equity algorithm compares the rate of patients under each racial and ethnic category in the metacohort to the hypothetical ideal rate of the same subgroup in an equitable RCT estimated from NHANES. Overall by leveraging the study cohort extraction, standardization, and equity algorithms, we can measure the equity gap in the guideline metacohorts and identify where more equitable clinical trial evidence is needed.