



Research Week 2021

The Quality of Clinical Social Determinants Data: A Systematic Review

Lily A. Cook, Nicole Weiskopf, Jonathan Sachs
cool@ohsu.edu

OHSU Department of Medical Informatics and Clinical Epidemiology

Keywords

Social determinants of health, data quality, secondary use, social informatics

Abstract

Objectives

Our objective was to conduct a review of literature on social determinants of health (SDoH) in clinical data in order to characterize data quality (DQ) issues specific to this area of research. To reach this objective, we conducted a realist review - a theory-driven, deductive, and systematic literature review.

Methods

We searched PubMed and Ovid MEDLINE databases for articles focused on social determinants of health that also included terms related to data quality. A snowball technique was used to identify additional works missed by the initial query. The results were then screened for relevance based on title and abstract. Articles were excluded if they were not in English, not original research, utilized datasets from institutions outside the US or Canada, or did not use a clinical data source such as an EHR or a registry.

Articles were appraised according to the elements of the Harmonized Data Quality Terminology and Framework, and categorized by the SDoH variable examined. This information was extracted and aggregated to identify patterns and themes.

Results

The most frequently explored issue was the plausibility (i.e., accuracy) of patients' race or ethnicity information (23 of 70 articles). Race/ethnicity was the social determinants variable with the largest body of DQ literature (39/70).

Geospatial data, used to link patient-level data to community-level datasets to incorporate variables unavailable in the EHR, was the next-most commonly represented (15/70). Relational conformance (i.e., linkage match rates) was the most highly explored data quality problem for geospatial data (10/15).

Also represented in the literature were environmental factors (5/70), insurance status (1/70), and occupation (1/70). Seven articles addressed SDoH generally without focusing on any specific variable.

Throughout all the literature, the most commonly represented issue was plausibility (30/70). A majority of the articles (46/70) noted the possibility of bias or validity issues being introduced into research by these problems, and 42 made research-based recommendations for improving DQ.

Significance

Interest among clinicians and researchers in social determinants of health has increased in recent years, driven by a recognition of social determinants of health as major contributors to health outcomes and by a desire to improve health equity. However, there are substantial and justifiable concerns regarding the quality of social determinants in clinical data. Without attention to the quality of this data it cannot be guaranteed that any research performed with it is providing valid or useful insights.

This research is unique because it brings together information about a variety of social determinants variables in order to examine the issues inherent to social informatics as a field of study. It is the first to summarize how specific issues of bias and validity are introduced into social determinants research, and to gather information about how the process of linking datasets has the potential to introduce data quality issues into research that relies on integrated datasets.