Reexamining Social Determinants of Health Data Collection in the COVID-19 Era

Samyukta Mullangi, MD, MBA
Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York.

Emeline M. Aviki, MD, MBA
Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York.

Dawn L. Hershman, MD, MS
Department of Medicine, Columbia University Medical Center, New York, New York.

Corresponding Author: Samyukta Mullangi, MD, MBA, Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10065 (smullangi13@gmail.com).

The COVID-19 pandemic put the most socially marginalized and vulnerable populations at further risk for physical, social, and financial harm, thereby exacerbating existing structural health care disparities. Yet, our understanding of the extent of these harms often comes from low-quality data sources—anecdotes, single-institution experiences, and analyses that use indirect socioeconomic data from the US Census about the area where a patient lives. The acuity and speed with which the pandemic wreaked havoc on society highlighted how we lack a sufficient data infrastructure to measure and respond to social determinants of health (SDOH).

Recently, our team used data from the American Society of Clinical Oncology’s (ASCO) COVID-19 in Oncology Registry to study the association of various factors, including demographic factors, disease-related factors, and SDOH, with cancer treatment delays among a cohort of patients across multiple sites who had all tested positive for SARS-CoV-2.1 As we had hypothesized, patient factors such as Black race, Hispanic ethnicity, multimorbidity, and timing of COVID-19 diagnosis were all independently associated with rates of treatment delay.1 To measure the effect of other types of SDOH data, we studied the association of 5 variables from the American Community Survey that were linked to the patient data by address—these were US Census tract-level racial and ethnic make-up, median household income, health insurance level, and education level, to allow for analysis about whether these SDOH were also associated with treatment delay.2 To our surprise, when controlling for patient-level variables, we found little association between these additional SDOH variables and the outcome of treatment delay. But in retrospect, we should not have been surprised at all.1

There is rising recognition of the effect of factors such as where our patients are born, raised, schooled, and work, on their health—collectively these are termed social determinants of health.3 Currently, SDOH data are most often being collected by entities such as public health agencies, the US Census Bureau, the Internal Revenue Service, and many nonprofits. These data are often incomplete—partly due to the limited and uneven regulations that govern data collection, and inconsistent use of data standards, which affect collective data integrity. The actual purveyors of health care, be they physicians, public health agency administrators, or insurers, frequently find themselves lacking the same information to pair with patients’ clinical data. During a year when patients experienced varying health outcomes by SDOH variables such as race, ethnicity, employment, and insurance status, the effect of this disconnected data architecture was much more acutely felt. But even if health care professionals were to pursue the imperfect linkage process to access SDOH data from these multiple sources, several challenges become apparent.

The first big challenge is that SDOH data garnered from these sources offers, at best, indirect inferences about what our patient cohort is likely experiencing. Though such data can enable users to conclude a certain disparity is driven by social factors, they say less about the exact proximal or distal cause of the disparity. Further, because the unit of measurement becomes more and more distant from the individual patient, it is increasingly prone to inaccuracies and misclassification biases, and mechanistically results in attenuation bias. One solution would be to direct the data collection process to health systems, and increasingly, standardized patient risk assessment protocols for screening and referral are being introduced to the health care delivery arena.4

Second, these data sources have their own built-in delay between collection and reporting, with even further delays to link given the data use agreements, anonymization, and institutional review board approvals that need to be put into place before linkages can occur. One recent novel study5 that linked cancer registry data with credit reports to study the association between a cancer diagnosis and risk of adverse financial events shared that the database linkage process took nearly 2 years to complete—time largely spent on establishing data use agreements between the cancer center and the credit agency. Even outside the context of a pandemic, such time delays dilute the effect of research findings, particularly in disparities research wherein the findings should ideally influence subsequent equity-oriented interventions such as referrals to community-based organizations. Further, as the experience of the pandemic highlighted, data sources do not capture data in motion. Although some SDOH variables such as race and ethnicity may be static, others such as employment, housing status, and social isolation may be quite dynamic.

But “better data” can be hard to define, particularly because different researchers will bring different needs to it. Two priorities that we believe would garner universal support, particularly in the wake of the pandemic, would be to secure more real-time data and more patient-level data. These are discrete, yet highly interrelated, and ideally immediately linked features. To achieve both, we believe the locus of data collection should reside at the health system level.

Data timeliness, a consumer-driven phenomenon, was ascendancy even prior to the pandemic. Customers, or patients, have grown accustomed to the convenience of real-time access to people and information.
Further, efforts at value-based care require data to be translated into actionable insights far sooner than in the encounter-based fee-for-service business model. Indeed, by replacing fee-for-service with episode-based payments, historic value-based payment models inherently incentivize health systems to address the negative externalities brought on by SDOH (ie, worse clinical outcomes and increased preventable health care utilization). New value-based models announced by the Center for Medicare and Medicaid Innovation, such as ACO REACH, are explicitly focusing on health equity as a priority, a necessary precondition to which is securing accurate data on population groups and social needs. Innovations in health information technology, including the adoption of common language standards to enable data interoperability, are increasingly making the collection of real-time data at the health system level more feasible. Architects of health system data need to also reduce regulations that limit researchers from analyzing this data and disseminating findings to ensure greater effect.

Securing patient-level data is a priority. Such data are critical to understanding the interplay between factors at one level (such as personal behaviors) while controlling for potential confounding at or the interaction with another level (such as neighborhood differences). Understanding the interaction between these different etiologic factors situated at different levels of analysis will enable interventions to be targeted with greater precision. Moving forward, we need to replace isolated snapshots of patient health with routine, prospectively collected, systematic, patient-reported assessments that automatically link to existing patient data. In future pandemics, patients will not only embrace, but will expect that their physicians engage in 2-way communication with them—both sharing critical safety information and collecting valuable data to inform population health.

Collecting more comprehensive and longitudinal SDOH data and making it concurrently available for use in both designing clinical care programs and informing research implies the creation of a learning health system. There also exists a responsibility that comes with collecting SDOH data, such as organized interventions or referrals for those who demonstrate essential needs that are not addressed. In addition to the technical challenges, other barriers to overcome include alignment among stakeholders and along operational time scales. An architectural design that can support clinical care, quality improvement, and research as integrated activities will require harmonizing terminology, a shared vision across stakeholders, strong partnerships, and funding, but the payoff will be well worth it.

**REFERENCES**