Mathematica
Progress Together

Health Research Brief

Author: Xindi (Cindy) Hu, ScD

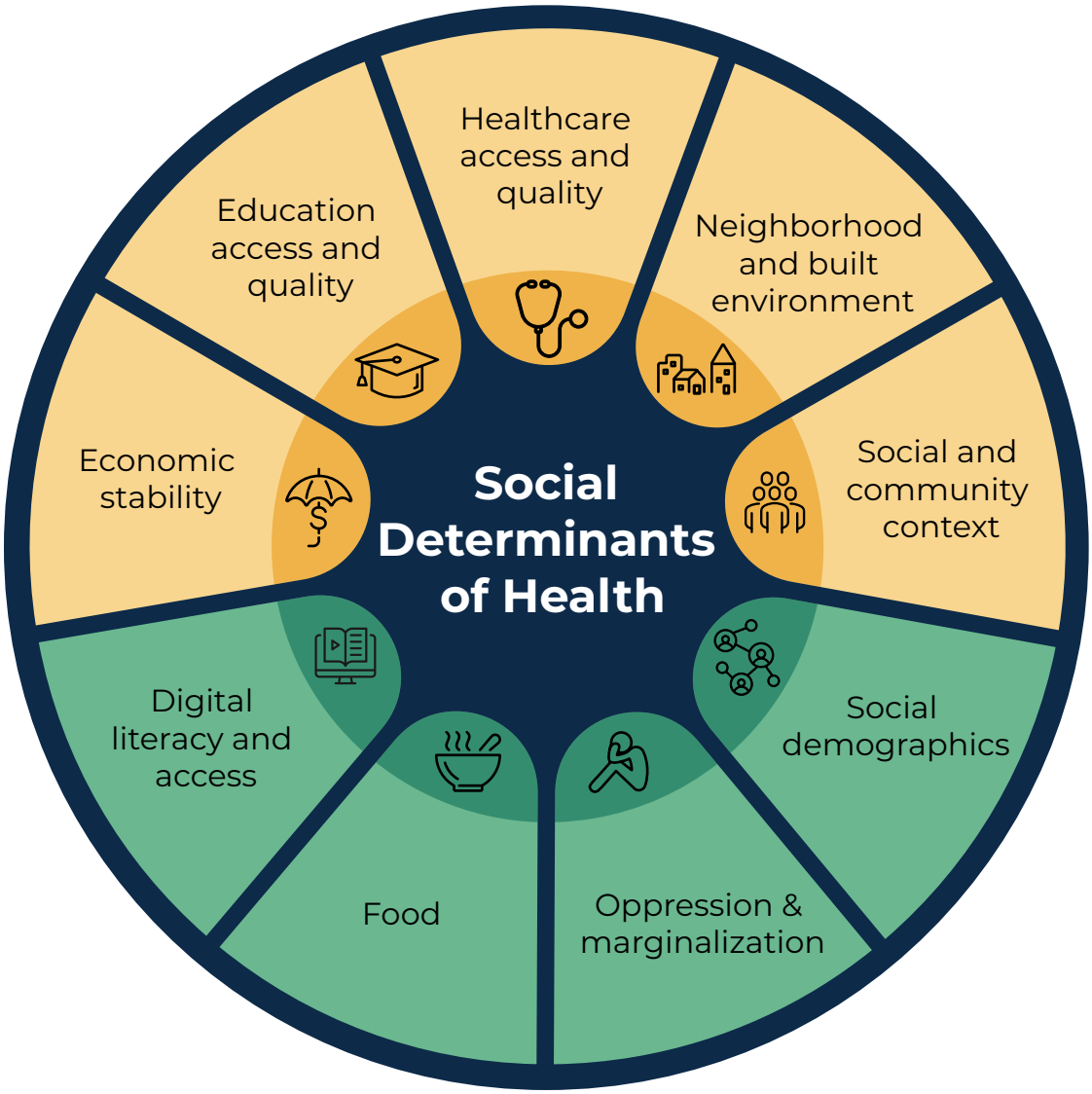
Five Considerations in Selecting Social Determinants of Health Data to Advance Health Equity

In the pursuit of health equity, data on social determinants of health (SDOH) have emerged as a crucial tool. The SDOH are defined as the conditions of where we are born, live, learn, work, play, worship, and age. They are shaped by structural, political, and economic forces, and contribute to persistent health disparities. Research shows that SDOH can account for 30 to 80 percent of the variation in health outcomes. When used properly, SDOH data can inform policies and interventions that address the root causes of health inequities. However, selecting the right SDOH data is a complex task that requires careful consideration. In this white paper, we explore the key factors in selecting and using SDOH data effectively to advance health equity.

1. Understanding the scope and relevance of SDOH data

When selecting SDOH data, it's essential to select the appropriate framework for your research question, which then helps identify the appropriate data and increases the interpretability of the results. One of the most widely adopted SDOH frameworks, developed by the Department of Health and Human Services, includes five domains: (1) economic stability, (2) education access and quality, (3) health care access and quality, (4) neighborhood and built environment, and (5) social and community context. Other organizations have expanded this framework (Exhibit 1). For example, the SDOH taxonomy from Kaiser Family Foundation added food, a sixth domain which includes both food security and access to healthy food options. Guided by Mathematica's Health Equity Framework, which uses a tree to represent the multilevel determinants of health and well-being, we propose adding other domains such as oppression and marginalization and physical and chemical environment. In addition, Canadian researchers developed a 13-category SDOH framework tailored to rural residents. Given the breadth and multifaceted nature of these factors, it's essential to select SDOH data that are directly relevant to a particular causal pathway to influence health outcomes. This is not an easy task—many SDOH indices combine measures into a single score, and different measures might influence health via different causal pathways. Therefore, we recommend consulting subject matter experts, including people with lived experience, in the area where the measure is being used. It is possible that researchers may only want to include a subset of domains, not all domains. Within each domain, there might be multiple indicators that researchers want to disaggregate further. For instance, when evaluating the Medi-Cal medically tailored meals pilot program, , Mathematica researchers compared health outcomes of program participants and nonparticipants living in similar areas, where similarity was defined by SDOH domains around food and economic stability. Because the mechanisms of SDOH that influence health may vary by outcome, the SDOH measure should be chosen according to its direct impact on the specific health outcome.

Exhibit 1. Core and amended domains in SDOH taxonomy

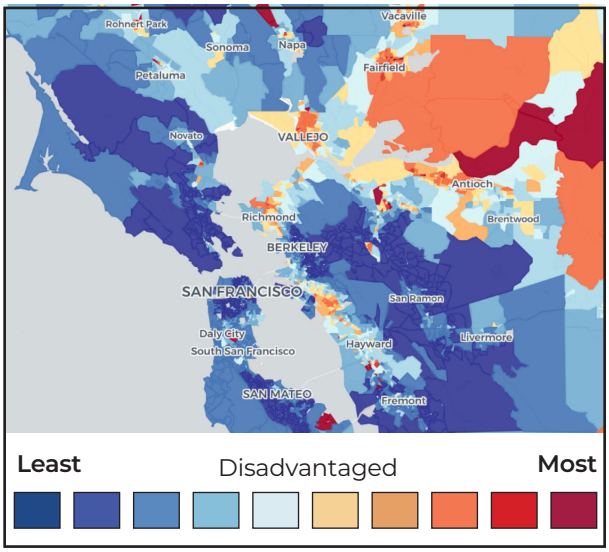


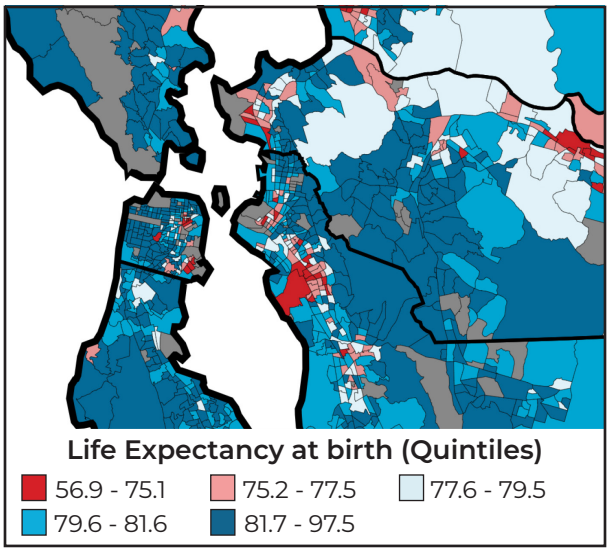
Note: the domains shaded in yellow are from the SDOH taxonomy developed by the Department of Health and Human Services

2. Differentiating between single indicator and composite indices

The choice between a single indicator (for example, the percentage of people with high school diplomas in a census tract) and a composite index (like the Area Deprivation Index [ADI] and Social Vulnerability Index [SVI]) will be affected by the specific research question and by whether a broad or detailed analysis is required. While indices can simplify complex data by synthesizing multiple SDOH variables into a single metric and make it easier to integrate with downstream data analytics, they might pose a challenge for interpretability and mask underlying disparities if not constructed or applied carefully. For example, issues such as a lack of standardization in input variables can lead to biases, as seen in the ADI's overweighting of median home value, which obscured disparities in high-cost-of-living areas (Exhibit 2). This issue was only discovered in mid-2023 after ADI had been used in federal and state healthcare policies for years. Since then, the Centers for Medicare and Medicaid Services has introduced a standardized ADI to reduce the rural-urban differences due to prices.

Exhibit 2. Northern California Area Deprivation Index from 2014 to 2018 with life expectancy from 2010 to 2015





Source: Azar et al, 2023, Health Affairs

3. Assessing data quality

The effectiveness of SDOH data in advancing health equity depends on its quality. High-quality SDOH data should be:

- **Accurate**—reflective of true social conditions and able to capture the experiences of the populations being studied
- **Statistically rigorous**—statistical methods (such as factor analysis) used for constructing indices must be clearly documented and sufficiently justified
- **Complete**—inclusive of all relevant variables and able to cover the entire population or community of interest
- **Granularity**—capable of measuring SDOH at the appropriate geographic level
- **Timely and sustainable**—reflective of current conditions and updated regularly to capture changes over time

SDOH data sources are diverse, varying in collection methods, standards, temporal and spatial resolution, and levels of detail. This diversity may lead to bias in reflecting the true social conditions. To address these challenges, the research community has employed multiple statistical methods to construct valid and reliable measures, including percentile ranking, principal component analysis, and multidimensional item response theory models. Each method has its advantages and disadvantages. When choosing an SDOH data source, researchers need to pay attention to the assumptions embedded in the statistical methods, as well as any evidence for validity, such as convergent validity and predictive validity.

When selecting SDOH data, there's a trade-off between including all relevant variables and covering the population of focus. Alongside nationally consistent sources like ADI, SVI, and County Health Rankings, some states have developed state-specific SDOH data. For instance, the California Healthy Places Index integrates 25 community-level characteristics, which is more comprehensive than the 16-17 indicators summarized by ADI and SVI. These include California-specific variables like voter turnout data from University of California, Berkeley, and environmental data from the California Environmental Protection Agency. While incorporating similar variables could enhance national data sets, researchers must consider the risk of incomplete coverage (see question 2 in Exhibit 3).

It's also crucial to consider whether the data source has a long-term maintenance plan (see question 4 in

Exhibit 3). During the COVID-19 pandemic, indices like the Minority Health SVI and the Pandemic Vulnerability Index were developed to aid community response. However, the sustainability of these indices is still uncertain. Sustainable data sources are vital for tracking progress, identifying trends, and making informed decisions over the long term.

4. Comparing advantages and limitations of individual- and community-level social risk data

When comparing individual-level and community-level social risk data, it's important to recognize the distinct advantages and limitations of each (see question 1 in Exhibit 3). Individual-level social risk data, typically collected during healthcare encounters, offers specific insights into personal social needs but is not widely available on a national scale. On the other hand, community-level social risk data offer broader national coverage and are more readily available. However, researchers need to carefully link these area-level data sources to individual-level health data. Aside from potential concerns of ecological fallacy (defined as an incorrect inference about individual-level relationships drawn from group-level data), the exact data linkage method also requires careful consideration (see questions 5 and 6 in Exhibit 3). The most used geographical identifiers include ZIP code (and by extension ZIP Code Tabulation Areas), county, or full address. However, despite its popularity, the ZIP code represents a mail route rather than any human behavior and is not ideal for spatial analysis. For example, in 2019, the 33139 ZIP code in Miami Beach, FL, had a median household income difference of more than $240,000 between the highest and lowest-earning census block groups. In addition, ZIP codes often do not align well with census geographic entities like block groups, census tracts, and counties. Alternatives to using ZIP codes include geocoding full addresses, collecting data in census geographic entities, and converting addresses to spatial indices (for example, hexagon and square grids). Spatial index and raster-based SDOH data have gained popularity recently in both public and private SDOH data sources. For example, NASA's Socioeconomic Data and Applications Center has developed gridded layers of Centers for Disease Control and Prevention SVI data at a spatial resolution of 1 kilometer. Similarly, the community risk scores developed by Socially Determined are available at 200 to 400-meter hexagons, depending on population density. Key advantages of raster-based SDOH data include high spatial resolution, scalability, and easy integration with user-defined geographical boundaries.

While health care studies typically focus on individual needs using data from social needs screenings during health care encounters, population health studies tend to prioritize community-level SDOH for its broader coverage, regardless of whether an individual comes into interaction with the health care system. Recent research highlights the importance of considering data on both individual- and community-level social risks. This dual focus is essential because the interaction between personal social challenges and broader community conditions can significantly affect health outcomes and together result in increased healthcare utilization. Focusing solely on community-level determinants without considering individual needs might overlook crucial aspects of health disparities. Conversely, focusing only on individual needs without context can lead to interventions that are disconnected from the broader social environment.

5. Ethical considerations and community engagement

Because SDOH data vary widely in their origins, quality, and potential impact on communities, ethical considerations are crucial in selecting these sources. Key questions include: Are the data collected with informed consent and cultural sensitivity? How is privacy protected? Are the data free of bias?

SDOH data come from a variety of sources, including the publicly available databases of state and federal agencies (Census, public health surveillance), insurance claims and electronic health records, and private sector entities that provide credit reporting, consumer purchasing,

and marketing data. The use of individual health-related social needs data could raise significant ethical concerns, especially when such data, initially intended for patient care, is repurposed for research without proper consent. In addition, the unintended consequences of using SDOH data, such as overlooking communities with less robust data collection or underrepresenting certain populations, can exacerbate health disparities rather than mitigate them. Biases in SDOH data—stemming from flawed collection methods, unbalanced population representation, or cultural insensitivity—can further skew research findings and lead to ineffective or harmful policy decisions, making it crucial to address these ethical issues carefully.

To mitigate these risks, it's crucial to engage communities in the data collection and analysis process actively through community *ground-truthing*, a process of verifying the accuracy and relevancy of data by comparing it with "on-the-ground" information. Communities are the experts on their own social conditions, and their input is invaluable in ensuring that data accurately reflect their experiences and needs. Engaging communities in selecting SDOH data fosters trust and ensures that the data used are not only theoretically sound but also practically valid and reflective of actual social determinants affecting health outcomes.

Applying SDOH Data to Drive Action

Exhibit 3 summarizes key considerations for selecting SDOH data. One of the first questions to ask is whether the focus is on individual-level social needs or community-level SDOH. After that's decided, we can ask ourselves other important questions including: What domains of SDOH are most relevant to the outcome? Do we need a summary index, or do we need subdomains? Is the method clearly documented? If multiple data sources are used in calculating the index, are they properly standardized? How important is being nationally consistent? Would a state-specific index be better suited? Are there plans to keep this index sustainable and updated over time? Is it easy to link to geographical identifiers in the health data? Are the data collected with ethical considerations and cultural sensitivity? The decision tree is meant for general guidance and users should consider specific project needs when selecting and using the SDOH data source.

Selecting and analyzing the right SDOH data informs actions that advance health equity. This requires moving beyond data collection and analysis to implementing evidence-based interventions that address the root causes of health disparities. By focusing on the relevance, quality, integration, and ethical use of the data, and by balancing individual-level social needs with community-level determinants, we can develop a deeper understanding of the social factors affecting health outcomes and take meaningful action to address them. Through collaborative efforts and a commitment to equity, SDOH data can serve as a powerful tool in the ongoing effort to reduce health disparities and promote better health for all.

Acknowledgments

I would like to express my sincere gratitude to Dr. Alex Bohl and Dr. Eli Michaels for their insightful comments and feedback on an earlier draft of this white paper.

Glossary

ICD: International Classification of Diseases

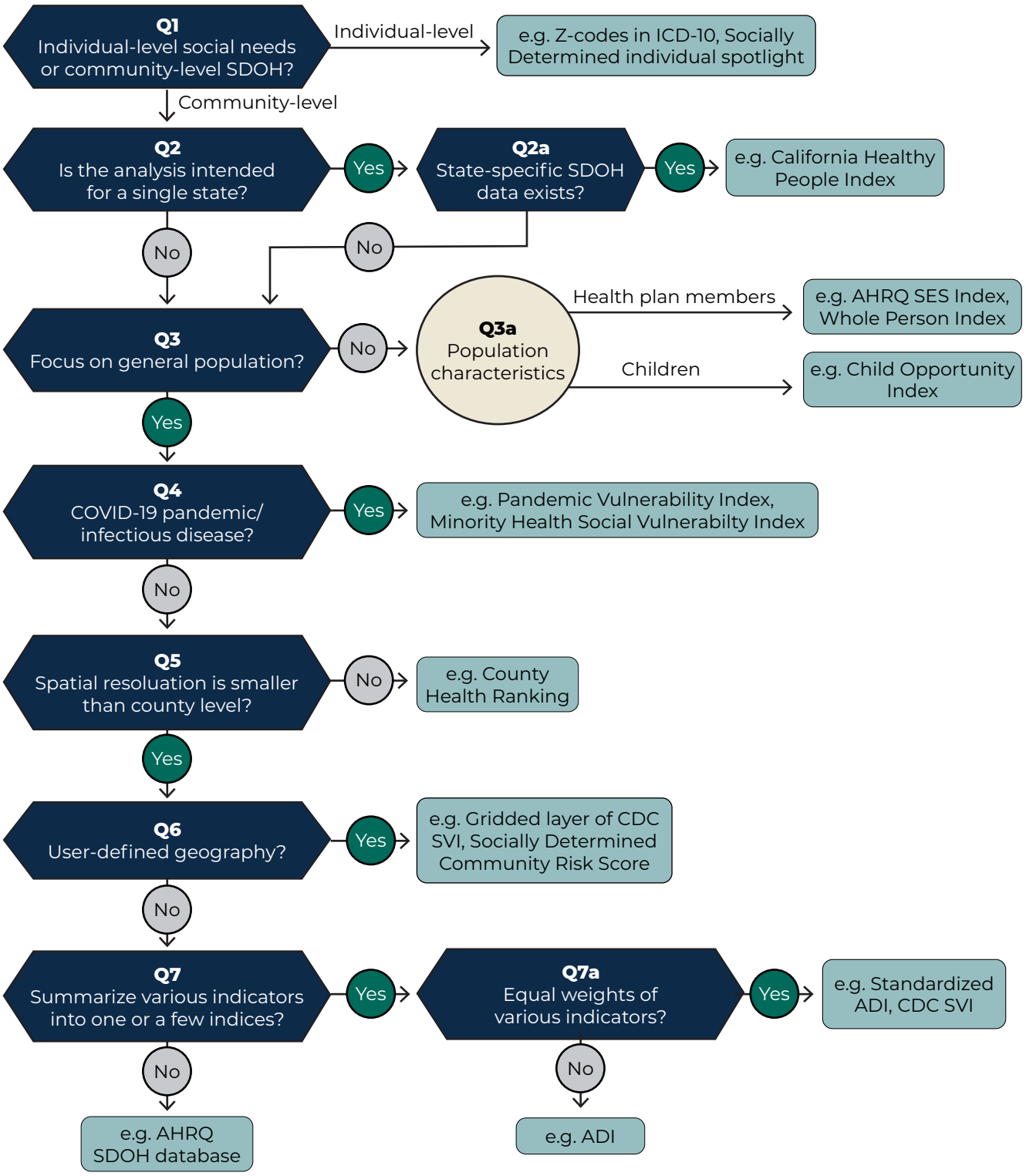
AHRQ: Agency for Healthcare Research and Quality

SES: Socioeconomic status

CDC SVI: Centers for Disease Control and Prevention's Social Vulnerability Index

ADI: Area Deprivation Index

Exhibit 3. Decision tree for selecting SDOH data source



Source: Mathematica developed based on project experiences between 2016 – 2023.