

Utilizing Publicly Available Community Data to Address Social Determinants of Health: A Compendium of Data Sources

INQUIRY: The Journal of Health Care
Organization, Provision, and Financing
Volume 60: 1–11
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00469580231152318
journals.sagepub.com/home/inq



Zoe Lindenfeld, B.S.¹ , José A. Pagán, PhD¹,
and Ji Eun Chang, PhD¹

Abstract

To compile a compendium of data sources representing different areas of social determinants of health (SDOH) in New York City. We conducted a PubMed search of the peer-reviewed and gray literature using the terms “social determinants of health” and “New York City,” with the Boolean operator “AND.” We then conducted a search of the “gray literature,” defined as sources outside of standard bibliographic databases, using similar terms. We extracted publicly available data sources containing NYC-based data. In defining SDOH, we used the framework outlined by the CDC’s Healthy People 2030, which uses a place-based framework to categorize 5 domains of SDOH: (1) healthcare access and quality; (2) education access and quality; (3) social and community context; (4) economic stability; and (5) neighborhood and built environment. We identified 29 datasets from the PubMed search, and 34 datasets from the gray literature, resulting in 63 datasets related to SDOH in NYC. Of these, 20 were available at the zip code level, 18 at the census tract-level, 12 at the community-district level, and 13 at the census block or specific address level. Community-level SDOH data are readily attainable from many public sources and can be linked with health data on local geographic-levels to assess the effect of social and community factors on individual health outcomes.

Keywords

data, social determinants of health, healthcare access, economic stability, health outcomes

What do we already know about this topic?

Community-level SDOH data represents a useful alternative to individual-level data to guide policymakers and providers in decision-making.

How does this topic contribute to the field?

We demonstrate the range of community data that can help identify SDOH, and that can be linked with health data to assess effects of social factors on individual health outcomes.

What are your research’s implications toward theory, practice, or policy?

The process employed here can prove useful to organize and assess SDOH data in any geographic area and across multiple policy-relevant dimensions.

Introduction

There is growing recognition that the conditions in which people are born, grow, work, live, and age shape health outcomes and equity. Understanding data on these social determinants of health (SDOH) is fundamental to improving population health and reducing health disparities, particularly for populations with high rates of chronic health conditions and deaths due to preventable diseases.^{1,2} In April 2022, the Biden Administration, in conjunction with the Department

of Health and Human Services, released statements on the Equity Action Plan, which aims to advance health equity for underserved and low-resourced populations.³ Along with maintaining a focus on SDOH, a key component of this plan is to advance the government’s ability to analyze disaggregated, demographic data to advance equitable health outcomes in communities across the U.S.⁴

Community-level SDOH data are oftentimes used to guide the allocation of resources and funds, developing new initiatives aimed at addressing health equity, and ensuring



that such services are reaching communities in greatest need.^{5,6} In the U.S, community-level SDOH data are collected by federal, state, and city-level government agencies through administering surveys, aggregating data from various public entities (eg, high school graduation rates, unemployment counts, subway, and bus usage rates), or as part of routine data collection procedures (eg, local tree counts, grocery store locations, environmental toxin measures).⁷ Using such community-level SDOH data can provide important insights to best address health disparities.

Community-level SDOH data are both highly interrelated and varied, with markers of SDOH including measures of income and wealth, education, work, political voice, social connections, environment, physical security (eg, crime), access to medical care, and residential environments. This multi-dimensional nature of SDOH makes it challenging to catalog and measure. Previous resources have coalesced data that encompasses a wide range of SDOH markers; at the federal level, the Agency for Healthcare Research and Quality (AHRQ) social determinants of health database links together SDOH data that are available nationwide at the zip-code level.⁷ At the city level, the Factors Affecting Communities and Enabling Targeted Services (FACETS) database links SDOH data available at the census-tract level in New York City.⁸ Such resources, and the SDOH data contained within them, have been used for various purposes, including for evaluation of government programs, research, and to develop health care and public health interventions.^{7,9} Yet, neither database uses a clear process for determining where to locate SDOH data and what type of SDOH data to include; additionally, both rely heavily on data stemming from the Census Bureau's annual American Community Survey (ACS), with less of a focus on locally-collected data. Incorporating a range of community-level data beyond what is included in the ACS is important, as previous studies have demonstrated the limits of relying on ACS data alone, which include mis-identifying or failing to identify persons in need.¹⁰

Given that SDOH measures come from a variety of sources, policymakers, and public health authorities may lack knowledge around what data sources are publicly available, at what geographic level, and what can be feasibly linked using geographic identifiers, such as zip-codes or community districts, within data sources. For example, linking SDOH data to health system data, such as patient-level EHR data, using zip-code level identifiers can remove the burden on health systems to collect patient-level SDOH data, as well as ensure that treatment decisions are tailored to

patients' needs.¹¹ As such, a compendium of SDOH data is an important resource to demonstrate the variety of data sources available at the local geographic level, which may be useful in guiding decision-making and resource allocation for populations with high numbers of social needs. To address the gap in knowledge surrounding available data, we used New York City (NYC) as an example and compiled a compendium of data sources representing different SDOH areas relevant to policymaking related to addressing population health challenges and achieving health equity. We demonstrate the range of community-level data that can be utilized to identify SDOH needs on a local scale, as well as our process for determining which data to include. The process employed here can prove useful to organize and assess SDOH data in any geographic area and across multiple policy-relevant dimensions.

Methods

We conducted an iterative and deductive search of the peer-reviewed literature and gray literature pertaining to SDOH data in NYC. The search was conducted as part of the AI4HealthyCities Health Equity Network project created by the Novartis Foundation. We first carried out a PubMed search on March 7th, 2022, using the search terms "social determinants of health" and "New York City" along with the Boolean operator "AND." There were no restrictions based on language, publication status, or year of publication. We then conducted a search of the "gray literature," defined as sources outside of these standard bibliographic databases to identify any data sources not utilized in the peer-reviewed literature. Gray literature sources included Google, Google Scholar, as well as the databases of federal, state, and city agencies known to have a role in either the provision of health or social services, or in performing routine data collection activities. We used search terms based on those in our academic literature review, tailored to the search functions of each website. We only included publicly available data sources, as well as data sources which contained NYC-based data, though sources which included NYC data as well as data for other locations were included as well.

In categorizing our data sources, we used the framework for SDOH outlined by the Centers for Disease Control and Prevention's (CDC) Healthy People 2030, which uses a place-based framework to outline 5 key domains of SDOH: (1) healthcare access and quality, which includes issues such as access to healthcare, access to primary care, health

¹New York University School of Global Public Health, New York, NY, USA

Received 27 October 2022; revised 9 December 2022; revised manuscript accepted 5 January 2023

Corresponding Author:

Zoe Lindenfeld, Department of Public Health Policy and Management, School of Global Public Health, New York University, 726 Broadway, New York, NY 10012, USA.

Email: zll213@nyu.edu

insurance coverage, and health literacy; (2) education access and quality, which includes issues such as high school graduation, higher education enrollment, language and literacy, and early childhood development; (3) social and community context, including issues such as community cohesion, discrimination, civic participation, workplace conditions, and incarceration; (4) economic stability, which includes poverty, employment, housing stability, and food stability; and (5) neighborhood and built environment, including issues such as access to transportation and healthy foods, air, and water quality, and neighborhood crime and violence.¹² Throughout our search process, we prioritized including data sources that are available at a smaller geographic level than the county level, such as community district, census tract or zip code, and aimed for wide representation of data sources within each SDOH category.

Results

The PubMed search yielded 126 results, from which we identified 29 publicly available datasets reporting SDOH data for New York City. We then identified an additional 34 publicly available datasets from gray literature sources, resulting in a total of 63 datasets related to SDOH outcomes in NYC. Of the identified datasets, 22 contained data related to the SDOH domain of healthcare access and quality, 18 to the domain of education access and quality, 22 to the domain of social and community context, 22 to the domain of economic stability, and 37 to the domain of neighborhood and built environment.

The data stemmed from a variety of sources, including federal departments, local health departments, police departments, and parks departments, among others, with 18 sources including data from the American Community Survey specifically. Among our data sources, 20 were available at the zip code-level, 18 at the census tract-level, 12 at the community district level, and 13 at the census block level or specific street address level, which could potentially provide more unique characteristics.

To assist in data linkage, there exists publicly available crosswalk files that link up the various local geographies in NYC, including the levels included in our findings. (NYC Crosswalk, 2022) The results of our search sorted by SDOH domain can be found in Table 1.

Discussion

There is increasing recognition that addressing SDOH can have a significant impact on population health and reducing health disparities. Infection rates and mortality during the COVID-19 pandemic differed by socioeconomic status and where people work and live. U.S. counties with higher death rates had higher proportions of Black residents, more uninsured adults, higher incarceration rates, more adults without a high school diploma, and more households without

internet.^{13,14} This has been particularly evident in NYC, in which there has been a strong association between socioeconomic status and health outcomes,¹⁵ as well as strong disparities in exposure density and disease outcomes at the neighborhood level.¹⁶⁻¹⁸ For example, in Yorkville, a neighborhood in Upper Manhattan with a 2019 median household income of \$136 560, a life expectancy of 85.9 years, and a population that was 72.5% White and 2.3% Black, 1 out of 509 people died from COVID-19. By contrast, in Brownsville, Brooklyn, which had a 2019 median income of \$31 880, a life expectancy of 75.1 years, and a population that was 68.4% Black and 3.8% White, 1 out of 139 people died of COVID-19.¹⁹⁻²¹ Similarly, Corona, Queens, in which 53.5% of residents identify as Hispanic and 35.3% of residents identify as Asian,²⁰ had the highest poverty rates in Queens as of 2019, with 27% of residents living in poverty,¹⁹ and had the highest reported number of confirmed COVID-19 cases as of May 2021, during the first wave of the pandemic in NYC.²² As such, it is imperative that research on health disparities must be informed equally by indicators collected within the clinic walls, as well as those that extend to the places where people live their daily lives.

In the U.S., a number of emerging initiatives aim to increase focus on SDOH data within decision-making, both within policymaking and in clinical settings. At the federal level, the Centers for Medicare & Medicaid Services launched the Accountable Health Communities model to systematically identify and address the social needs of Medicare and Medicaid beneficiaries through linkages to community-based resources. On a state level, there have been several initiatives which collect data on Medicaid enrollees' housing status and provide housing support or employment services to those with unstable housing.¹ Within medical settings, SDOH screening tools gather information related to a patient's SDOH needs in the context of a clinical visit. Examples include social needs screening tools designed for providers to engage patients in assessing SDOH,^{23,24} as well as International Classification of Diseases, Tenth Revision (ICD-10) Z codes, a set of 24 diagnostic codes used for tracking SDOH within electronic health record systems (EHRs).²⁵

However, many efforts to combine data entail manually collecting and coding SDOH data from individual enrollees—a laborious endeavor that is done sporadically at best. Within clinical settings, although providers may benefit from having knowledge of a patient's social needs, collecting SDOH data in the context of a medical visit may be time-consuming and infeasible. For example, estimates of Z-code collection suggest that utilization is low, with studies finding that only about 0.96% to 2.03% of patients have a Z-code included in their records.²⁶⁻²⁸ As such, having access to previously collected data pertaining to patients' community-level SDOH may reduce these barriers and help guide providers' decision-making. Additionally, a recent initiative from the National Institutes of Health, the All of Us Research Program, is a

Table 1. New York City Data by SDOH Domain.

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain		Description of data included
					(1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	
American Community Survey*	Census tract and zip code	X	2005-2020	Annually	#1, 2, 3, 4, 5	X	Source of statistics about socioeconomic and housing characteristics; covers a wide range of social, economic, housing and demographic topics
NYC DOHMH COVID-19 Data*	Zip code		2020-2022	Daily	#1		COVID-19 case and death rates in New York City
NYC Government Environment and Health Data Portal*	Community district		2003-2020	Varies	#1, 5		Data on how the environment affects health related to outdoor air and weather, built environment, pests and pesticide use, food and drink, environmental susceptibility, health outcomes, behavior, population characteristics, and vulnerability
United States Census Bureau*	Census tract and zip code	X	1990-2020	Every 10 year	#1, 2, 3, 4, 5	X	Demographic, economic and population data from the U.S Census
The United States Department of Agriculture's Food Environment Atlas*	FIPS code		2011-2020	Every 2 year	#4, 5	X	Food environment factors—such as store/restaurant proximity, food prices, food and nutrition assistance programs, and community characteristics
New York City Fire Department Housing Vacancy Data*	Street address		2010-2018	Every 3 year	#3		List of ordered vacated buildings declared unsafe and off-limits
New York Statewide Planning and Research Cooperative System*	Zip code		1982-2022	Every 6 month	#1		Patient level detail on patient characteristics, diagnoses and treatments, services, and charges for each hospital inpatient stay and outpatient (ambulatory surgery, emergency department, and outpatient services) visit; and each ambulatory surgery and outpatient services visit to a hospital extension clinic and diagnostic and treatment center licensed to provide ambulatory surgery services.
DOHMH Live Birth Records*	Community district		2000-2016	Annually	#1		All birth and deaths in NYC, including the number and rate of live births by mother and baby demographics, birth outcomes
GDC Environmental Public Health Data*	Census tract		1991-2016	Varies	#5	X	Data and information on environments and hazards, health effects, and population health.
NYC Complaint Data*	Street address		2006-2021	Annually	#5		All valid felony, misdemeanor, and violation crimes reported to the New York City Police Department

(continued)

Table 1. (continued)

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain (1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	Description of data included
NYC Facilities Database*	Street address		2021	Annually	#1, 2, 5		Captures the locations and descriptions of public and private facilities ranging from the provision of social services, recreation, education, to solid waste management.
NYC Subway Stations Data, NYC Open Data*	Location coordinates		2013-2019	Daily	#5		NYC subway station locations
NYC Tree Count Data*	Street address		1995, 2005, 2015	Unknown	#5		Number of trees, tree species, diameter, and perception of health
Metropolitan Authority New York City transit data*	Location coordinates		2021	Daily	#5		The number of MetroCard swipes made each week for the 472 subway stations in the NYC subway system
CDC Places*	Census tract and zip code	X	2016-2019	Unknown	#1	X	Model-based population-level analysis and community estimates on health-related outcomes
FACETS dataset*	Census tract	X	2016-2018	Unknown	#1, 2, 3, 4, 5		Architecture that incorporates open data related to SDH into a single dataset mapped at the census-tract level for New York City
Land Use Regression model obtained from the Center for Air, Climate and Energy Solutions*	Census tract and census block group		1979-2017	Unknown	#5	X	Provide estimates of outdoor concentrations for 7 pollutants (4 gases: O ₃ , CO, SO ₂ , NO ₂ ; 3 aerosols: PM ₁₀ , PM _{2.5} , PN [particle number])
Number of tobacco retailers per census tract in NYC*	Census tract		2020-2022	Annually	#5		Businesses/individuals holding a DCA license
National Highway Safety Administration FARS Data*	County		1982-2022	Annually	#5	X	Nationwide census providing yearly data regarding fatal injuries suffered in motor vehicle traffic crashes
NYU Furman Center*	Community district, sub-borough districts, congressional district, city council district	X	2011-2019	Annually	#2, 3, 4, 5		Housing and demographic trends in NYC neighborhoods
NY State Board of Elections Data*	Community district		1996-2022	Annually	#3		Data on voter enrollment

(continued)

Table 1. (continued)

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain (1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	Description of data included
CDC Social Vulnerability Index*	Census tract	X	2000-2022	Unknown	#1, 2, 3, 4, 5	X	Includes 15 U.S. census variables, including poverty, lack of vehicle access, and crowded housing, that are grouped into 4 related themes, including socioeconomic status; household composition and disability; minority status and language; and housing type and transportation.
NYC DOHMH HIV Surveillance Registry*	Zip code		2010-2013	Annually	#1		Data on new diagnoses of HIV and AIDS in NYC; reported cases and case rates (per 100,000 population) are stratified by age group, race/ethnicity
NYC Police Department Stop, Question, and Frisk (SQF) data*	Street address		2002-2021	Unknown	#3, 5		Data recorded by the police officer after a stop such as the age of the person stopped, if a person was frisked, if there was a weapon or firearm recovered, if physical force was used, and the exact location of the stop within the precinct
United States Postal Service/ Department of Housing and Urban Development Data*	Zip code		2005-2021	Annually	#3	X	Aggregate vacancy and no-stat counts of residential and business addresses that are collected by postal workers
New York City Department of Education FitnessGram records*	Census tract		2005-2022	Annually	#1, 2, 5		Data on student height, weight, fitness level, strength, endurance, and flexibility
New York State Public Access Cancer Epidemiology Database*	County, NYC neighborhood		1995-2022	Annually	#1		Tabulated cancer incidence and mortality statistics by race/ethnicity, gender, county, cancer site, and year of diagnosis.
NYC Open Data-Bus Stop Shelters*	Street address		2017-2021	Annually	#5		Location of bus stop shelters in NYC
NYC Community Health Survey*	Zip code		2002-2020	Annually	#1		Data on the health and risk behaviors of New Yorkers
Data2goNYC**	Community district, census tract	X	2014-2020	Unknown	#1, 2, 3, 4, 5		Federal, state, and city data on neighborhood assets and challenges
AHRQ Social Determinants of Health Database**	Zip code, county	X	2011-2020	Annually	#1, 2, 3, 4, 5	X	Variables correspond to 5 key SDOH domains: social context (eg, age, race/ethnicity, veteran status), economic context (eg, income, unemployment rate), education, physical infrastructure (eg, housing, crime, transportation), and healthcare context (eg, health insurance).

(continued)

Table 1. (continued)

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain		Description of data included
					(1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	
HRSA Area Deprivation Index**	Census block group	X	2016-2019	Unknown	#2, 3, 4, 5	X	Based on a measure created by the Health Resources & Services Administration (HRSA) over 3 decades ago, and has since been refined, adapted, and validated to the Census Block Group neighborhood level
County Health Ranking and Roadmaps**	County, zip	X	2010-2021	Annually	#1, 2, 3, 4, 5	X	Provides data, evidence, guidance, and examples to build awareness of the multiple factors that influence health and health equity
USDA Food Access Research Atlas**	Census tract		2006-2019	Every 4 year	#4, 5	X	Overview of food access indicators for low-income and other census tracts using different measures of supermarket accessibility
USDA SNAP Distribution**	Congressional district		1998-2021	Annually	#4, 5	X	Aggregate data on individuals and families receiving SNAP assistance
NYC Farmers Markets**	Street address, community district		2009-2020	Every 2 week	#3, 5		List of farmers markets
NYC School Attendance and Enrollment Statistics by District**	Community district		2011-2021	Annually	#2		Department of Education attendance and enrollment statistics
School District Breakdowns NYC**	Community district		2011-2019	Annually	#2		Demographic Statistics broken down by school districts (race, gender, citizenship status, % receives public assistance)
Area Health Resources File**	County	X	2018-2021	Annually	#1	X	Data on health care professions, health facilities, population characteristics, economics, health professions training, hospital use, hospital expenditures, and environment
Community Resilience Estimates**	Census tract	X	2018-2019	Unknown	#3, 4	X	Metric for how at-risk every neighborhood in the United States is to the impacts of disasters, including COVID-19
Local Area Unemployment Statistics**	County	X	1976-2021	Annually	#3, 4	X	Data on unemployment rates by month and 12-month net changes.
Local Affordability Index**	Census tract	X	2014-2019	Every 5 year	#3, 4	X	Estimates of household housing and transportation costs at the neighborhood-level along with constituent data on the built environment and demographic characteristics

(continued)

Table 1. (continued)

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain (1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	Description of data included
Social Indicators Report**	Census tract	X	2021	Annually	#1, 2, 3, 4, 5	X	Data represents current conditions in our society in a number of domains (eg, Education, Health) and for a variety of populations (eg, Citywide, by Borough, Race/Ethnicity, Gender)
Opioid Environment Policy Scan**	Census tract, zip code	X	2021	Annually	#1, 3, 5	X	Data warehouse exploring the multi-dimensional risk environment, from neighborhoods to states, impacting opioid use and health outcomes
USCB—Zip Code Business Patterns**	Zip code		1994-2019	Annually	#5	X	Economic data by industry including the number of establishments, employment, first quarter payroll, and annual payroll
EPA, Safe Drinking Water Information System**	City, county		2013-2021	Annually	#5	X	Contains information about public water systems and their violations of EPA's drinking water regulations
Robert Graham Center- Social Deprivation Index**	Census tract, zip code, Primary Care Service Area	X	2008-2015	Unknown	#1, 2, 3, 4, 5	X	Quantifies levels of disadvantage across small areas, evaluate their associations with health outcomes, and address health inequities
NYC Park Crime Statistics**	Borough		2014-2021	Annually	#5		Reported major felony crimes that have occurred within New York City parks
Family Violence Related Snapshots: New York City Community Board Districts**	Community district		2018-2021	Annually	#5		Count data for the number of family-related domestic incident reports, family-related felony assaults, domestic violence related felony assaults, family-related rapes and domestic violence related rapes
Internet Master Plan: Adoption and Infrastructure Data by Neighborhood**	Zip code, neighborhood		2019-2021	Annually	#2, 3, 4		Key indicators of broadband adoption, service and infrastructure
Behavioral Risk Factor Surveillance System (BRFSS)**	Zip code		2002-2021	Annually	#1	X	Data on modifiable risk behaviors and other factors contributing to the leading causes of morbidity and mortality
Banked and unBanked in NYC**	Neighborhood, FIPs		2011-2013	Unknown	#4		Percentage of individuals without banking accounts by NYC neighborhood

(continued)

Table 1. (continued)

Data source	Smallest geographic levels	Contains mainly ACS data	Years available	Update frequency	SDOH domain (1- Healthcare, 2- Education, 3- Social/Community, 4- Economic, 5- Neighborhood)	Contains non-NYC data	Description of data included
CityHealth Dashboard**	Zip code		2020-2021	Annually	#1, 2, 3, 4, 5	X	35+ measures of health, the factors that shape health, and drivers of health equity for 750+ U.S. cities
NYC Open Data Childcare Centers**	Zip code		2019-2022	Daily	#2		Data on capacity, and infractions in Childcare centers across NYC
Schools NYPD Crime Data Report**	Street address, community district		2017-2018	Unknown	#2		Data on crimes reported in schools
NYC 311 Data**	Zip code, address, geo coordinates		2010-2022	Daily	#3, 5		Data on 311 service requests by individual report
Built Environment and Health Research Group Neighborhood Walkability**	Census tract	X	1990-2015	Unknown	#5		Index measuring the walkability of neighborhoods
NYC ACS Total Admissions to Secure Juvenile Detention**	Zip code		2011-2022	Annually	#5		Total Admissions to Secure Detention by Age, Gender, and Race
NYC ACS Child Abuse/Neglect Report**	Community District		2016-2021	Annually	#4		Number of investigations and indication rate
NYC ACS Youth in Foster Care**	Community District		2019-2021	Annually	#4		The number of youths in foster care
NYC Open Data Green Spaces**	Address		2018-2022	Annually	#5		Identifies City property under the jurisdiction of NYC Parks
NYC Open Data Athletic Facilities**	Address		2018-2022	Annually	#5		Directory of Department of Parks and Recreation (DPR) athletic facilities
Prison Policy Initiative Number of People in Prison**	Zip Code		2010-2020	Every 10 year	#5		Imprisonment rate and aggregate number of people imprisoned

*Identified in PubMed Search.

**Identified in gray literature.

diverse collection of health data collected from study participants, including clinical indicators and medical conditions.²⁹ While the database does include a limited number of individual-level survey items pertaining to social factors, researchers utilizing the All of Us data, or any EHR data, may benefit from the inclusion of community-level SDOH data to inform their analyses and conclusions.

There are 4 main limitations to this paper. First, while the included data sources report health and social outcomes related to ethnic and racial minorities, there is a dearth of data illustrating SDOH for sexual and gender minority populations. Given that approximately 1 million adults in New York State identify as LGBTQ+,³⁰ this is a significant limitation to the data, and future efforts to collect data should incorporate measures related to sexual and gender identity to reflect the presence of this population. Second, there are a number of measures in our compendium that report important variables and concepts but exists in a format that requires significant data cleaning and reshaping, and that hinders ease of use. This includes data on crime (NYC Park Crime Statistics, Family Violence Related Snapshots, Schools NYPD Crime Data Report), as well as data on the availability and quality of childcare centers (NYC Childcare Centers). Third, the reported datasets varied in their timeliness of updates, which may limit their usefulness when linked to health system data, particularly given the rapidly changing landscape of NYC neighborhoods.³¹ Lastly, as the focus of our paper was NYC, we did not extract from or report datasets that did not include data specific to NYC. Future research to aggregate community-level SDOH data should focus on a wider range of cities, to better approximate the range of SDOH data that may be relevant for research or policymaking.

Despite these limitations, our results offer important enhancements on previous efforts to aggregate SDOH data, by incorporating a wider variety of locally-collected and NYC-specific data, that includes variables not reported in prior collections of SDOH data. In particular, this includes variables related to crime, school quality, neighborhood cohesion, public transit, and green spaces. Furthermore, we offer a general process for identifying, connecting, and interpreting the usefulness of SDOH data sources, which may be useful for researchers wishing to adapt this approach to alternate settings.

Conclusion

Community-level SDOH data can be obtained from a wide variety of public sources and can be linked with other data sources (eg, electronic health records data) based on local geographic identifiers. In particular, the identified data sources can be utilized for a variety of purposes, including policymaking, health services research, clinical decision making, and community resource allocation, to better address the range of social and community-wide factors that negatively impact individual health outcomes in an evidence-based manner.

Decision makers and researchers wishing to use such data in efforts to improve health should ensure wide representation of social determinants to enable the delivery of interventions that are tailored to the specific needs of populations.

Acknowledgments

We thank our colleagues Dr. Ann Aerts, Sarah Des Rosiers, and Johannes Boch for their expertise and insights in reviewing this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Novartis Foundation

Ethics and Informed Consent Statement

Our study of publicly available data sources did not require an ethical board approval.

ORCID iD

Zoe Lindenfeld  <https://orcid.org/0000-0002-5456-7665>

References

1. Artiga S, Hinton A. Beyond health care: the role of social determinants in promoting health and health equity. Kaiser Family Foundation. May 10, 2018. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/> (accessed July 17, 2022).
2. Andermann A. Taking action on the social determinants of health in clinical practice: a framework for health professionals. *CMAJ*. 2016;188(17-18):E474-E483.
3. United States Department of Health and Human Services (HHS). HHS statements on new plan to advance equity in the delivery of health and human services. 2022. <https://www.hhs.gov/about/news/2022/04/14/hhs-statements-on-new-plan-advance-equity-delivery-health-human-services.html> (accessed July 17, 2022).
4. The White House. Advancing equity and racial justice through the Federal government. 2022. <https://www.whitehouse.gov/equity/#health-equity> (accessed July 17, 2022).
5. Centers for Disease Control and Prevention. CDC programs addressing social determinants of health. 2021. <https://www.cdc.gov/socialdeterminants/cdcprograms/index.htm> (accessed July 17, 2022).
6. Gottlieb LM, Francis DE, Beck AF. Uses and misuses of patient- and neighborhood-level social determinants of health data. *Perm J*. 2018;22:18-078.
7. Agency for Healthcare Research and Quality. Social determinants of health database (Beta Version). 2021. <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html> (accessed June 1, 2022).

8. Cantor MN, Chandras R, Pulgarin C. FACETS: using open data to measure community social determinants of health. *J Am Med Inform Assoc*. 2018;25(4):419-422.
9. OPTUM. Social determinants: completing the health data picture. 2019. https://cdn-aem.optum.com/content/dam/optum3/optum/en/resources/articles-blog-posts/WF1213961_OptumIQ%20NTD%20article_SDOH.pdf (accessed July 17, 2022).
10. Cottrell EK, Hendricks M, Dambrun K, et al. Comparison of community-level and patient-level social risk data in a network of community health centers. *JAMA network open*. 2020;3(10):e2016852.
11. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. *Health Aff*. 2018;37(4):585-590.
12. Centers for Disease Control and Prevention. About social determinants of health. 2021. <https://www.cdc.gov/social-determinants/about.html> (accessed June 1, 2022).
13. Dalsania AK, Fastiggi MJ, Kahlam A, et al. The relationship between social determinants of health and racial disparities in COVID-19 mortality. *J Racial Ethn Health Disparities*. 2022;9(1):288-295.
14. Khazanchi R, Beiter ER, Gondi S, Beckman AL, Bilinski A, Ganguli I. County-level association of social vulnerability with COVID-19 cases and deaths in the USA. *J Gen Intern Med*. 2020;35(9):2784-2787.
15. Zhang Y, Khullar D, Wang F, et al. Socioeconomic variation in characteristics, outcomes, and healthcare utilization of COVID-19 patients in New York City. *PLoS One*. 2021;16(7):e0255171.
16. Hong B, Bonczak BJ, Gupta A, Thorpe LE, Kontokosta CE. Exposure density and neighborhood disparities in COVID-19 infection risk. *Proc Natl Acad Sci U S A*. 2021;118(13):e2021258118.
17. Carrión D, Colicino E, Pedretti NF, et al. Neighborhood-level disparities and subway utilization during the COVID-19 pandemic in New York City. *Nat Commun*. 2021;12(1):3692-3700.
18. Carrión D, Colicino E, Pedretti NF, et al. Assessing capacity to social distance and neighborhood-level health disparities during the COVID-19 pandemic. *MedRxiv*. Preprint. Posted online June 4, 2020. doi:10.1101/2020.06.02.20120790
19. NYC Health. Neighborhood data profiles. COVID-19: data. 2022. <https://www1.nyc.gov/site/doh/covid/covid-19-data-neighborhoods.page> (accessed June 15, 2022).
20. NYU Furman Center. New York neighborhood data profiles. 2022. <https://furmancenter.org/neighborhoods> (accessed June 15, 2022).
21. NYC Health. New York city community health profiles. 2022. <https://www1.nyc.gov/site/doh/data/data-publications/profiles.page> (accessed June 15, 2022).
22. The New York Times. New York city coronavirus map and case count. May 25, 2021. <https://www.nytimes.com/interactive/2020/nyregion/new-york-city-coronavirus-cases.html#zipcode> (accessed June 15, 2022).
23. National Association of Community Health Centers (NACHC). PRAPARE: Protocol for responding to and assessing patient assets, risks, and experiences. September 2, 2016. http://www.nachc.org/wp-content/uploads/2018/05/PRAPARE_One_Pager_Sept_2016.pdf (accessed July 17, 2022).
24. Centers for Medicare and Medicaid Services (CMS). The accountable health communities health-related social needs screening tool. Center for Medicare and Medicaid Innovation. 2021. <https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf> (accessed July 17, 2022).
25. Friedman NL, Banegas MP. Toward addressing social determinants of health: a health care system strategy. *Perm J*. 2018;22:18-095.
26. Weeks WB, Cao SY, Lester CM, Weinstein JN, Morden NE. Use of Z-codes to record social determinants of health among fee-for-service Medicare beneficiaries in 2017. *J Gen Intern Med*. 2020;35(3):952-955.
27. Guo Y, Chen Z, Xu K, et al. International Classification of diseases, tenth revision, clinical modification social determinants of health codes are poorly used in electronic health records. *Medicine*. 2020;99(52):e23818.
28. Truong HP, Luke AA, Hammond G, Wadhwa RK, Reidhead M, Joynt Maddox KE. Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016-2017. *Med Care*. 2020;58(12):1037-1043.
29. National Institutes of Health (NIH). All of Us research program. 2022. https://docs.google.com/document/d/1MaKk_awPaI5pT2GE5oJFFXDPFRARIK7kCIBrCgAtw0/edit (accessed August 1, 2022).
30. New York State Department of Health. Sexual Orientation and Gender Identity: Demographics and Health Indicators. BRFSS Brief Number 2022-16. June 6, 2022. https://www.health.ny.gov/statistics/brfss/reports/docs/2022-16_brfss_sogi.pdf (accessed August 1, 2022).
31. New York University Center for Urban Science + Progress. See how hard gentrification has hit your NYC neighborhood. April 21, 2019. <https://cusp.nyu.edu/news/gentrification-nyc-neighborhood/> (accessed August 1, 2022).