COVID-19 and the Causal Relationship between the Social Determinants of Health and Health Disparities

Leigh Wedenoja

Introduction

The COVID-19 pandemic has disproportionately affected Black and Hispanic Americans compared to their white peers. Adjusting for age, Black and Hispanic Americans have both a higher likelihood of contracting COVID-19 and a higher likelihood of dying from the illness (Artiga et al., 2021; CDC, 2020). The disparate impact of COVID-19 on Black and Hispanic communities is only one of many health disparities. Black and Hispanic Americans have lower life expectancy and higher incidents of nearly every disease (Wong et al., 2002).

Historically, research into the causes of these health disparities has focused on immediately causal or "downstream" factors including health behaviors and medical interventions. Over the past 20 years public health researchers and practitioners have increasingly recognized the role of the social determinants of health in maintaining and exacerbating disparities in health outcomes between racial and ethnic groups. These "upstream" factors include the situations in which people live, work, and study. Examples likely to directly affect COVID-19 include living in crowded housing and working in a job that requires close proximity to others. Other factors, like educational attainment (Zimmerman and Woolf, 2014), may impact the probability of COVID-19 transmission through intermediate causes. More educated workers are more likely to be able to work from home, have health insurance, and demonstrate greater health literacy (Schillinger, 2020).

The relationship between disparities in the social determinants of health (SDOH) and disparities in health outcomes is extremely complicated. Measures of SDOH such as educational attainment, income, health care access and affordability, employment, and neighborhood are inherently correlated and may affect health outcomes through more than one causal pathway. Further, causality could move in the opposite direction. People with disabilities or chronic illnesses may have limited job opportunities, less access to education, and lower incomes leading poor health to drive lower SDOH measures. It is extremely difficult to determine the causal effect of upstream SDOH on health outcomes and health disparities because experimentally assigning those factors is impossible, unethical, or both. In observational studies, strong assumptions are required to infer any causal relationship from a correlational estimate.

The COVID-19 pandemic presents an opportunity to estimate the causal relationship between upstream SDOH and health disparities, specifically disparities in COVID-19 incidence and mortality. Causal inference tools common in econometrics and statistics can be employed to exploit the "as good as random" or "quasi-experimental" variation in geography and timing of COVID's initial spread and surges as well as differences in the timing of mitigation measures and vaccine roll outs. These methods permit causal inference using a set of statistical assumptions.

COVID-19 also fundamentally altered the way people lived, worked, and studied which are SDOH factors and may impact other health disparities in the long term.

In order to be both statistically sound and socially valuable, any studies using COVID-19 variation to examine the link between SDOH and health disparities will require high quality data with measures of COVID-19 outcomes, SDOH, and documentation of the variation to be exploited. There are two types of variations that are likely valuable in the study of SDOH and COVID-19: pre-existing variation in the SDOH and geographic variation in COVID-19 prevalence and policies.

The rest of this chapter will proceed as follows. First, I discuss the complicated statistical relationship between upstream SDOH and health disparities. Second, I present pre-pandemic data for a subset of SDOH measures that may be causally related to disparities in COVID-19 transmission and mortality. Finally, I discuss examples of statistical and econometric methodologies and data that can provide insight into the causal relationship between SDOH and health disparities in the context of COVID-19.

The Complex Causal Relationship between the Social Determinants of Health and Health Disparities

The National Institute on Minority Health and Health Disparities defines a health disparity as "a health difference that adversely affects disadvantaged populations, based on one or more of" several outcomes including "higher incidence and/or prevalence and earlier onset of disease" and "premature and/or excessive mortality from diseases where the population rates differ" among others (NIH, 2021). By this definition, health disparities can exist and be measured between any relatively disadvantaged group and the relatively advantaged group. Here I focus specifically on the disparities in health for Black and Hispanic Americans compared to white Americans.

Since the early 2000s research into the causes of racial and ethnic health disparities has increasingly focused on the role of the social determinants of health (NICHD, 2000) resulting in a substantial increase in the number of organizations that address SDOH (Koh et al., 2010), SDOH research, and funding for SDOH research (Kneipp et al., 2018). These literature reviews focus on the complex causal interaction between upstream SDOH, downstream health behaviors, and health disparities (Braveman et al., 2011). Understanding and mapping the mechanisms through which upstream and downstream factors are related is crucial in identifying effective interventions for reducing health disparities (Alvidrez et al., 2019). Studying this complex process requires high quality data with accurate measurement of both contributing SDOH and resultant health disparities and advanced statistical methods (Penman-Aguilar et al., 2016).

The Department of Health and Human Services identifies five key areas of SDOH: Economic Stability, Education Access and Quality, Health Care Access and Quality, Neighborhood and Built

Environment, and Social and Community Context. Each of these five categories contain overlapping indicators that affect each other and downstream health behaviors.

Causal estimation of SDOH on health disparities is difficult compared to measuring the effects of health behaviors or public health interventions. The causal effect of an intervention, such as a smoking cessation program, can be identified through a randomized control trial (RCT). In the case of the SDOH, RCTs are often impossible or unethical. It is both impossible to randomly assign education and the relationship between education and health disparities is complex. Education could directly affect health through health literacy or it could affect health through higher income and better access to preventative medical care.

Suggestive Evidence for the Role of SDOH in COVID-19 Health Disparities

There is substantial evidence suggesting that the higher rates of COVID-19 incidence and mortality for Black and Hispanic Americans is driven by differences in SDOH that predated the pandemic (Singu et al., 2020; Burström and Tao, 2020; Turner-Musa et al., 2020; Abrams and Szefler, 2020). I present data on a subset of SDOH measures likely to be linked to COVID-19 transmission and mortality. This data analysis is partially adapted from the Rockefeller Institute report "Measuring Disparities in the Social Determinants of Health in Relation to the COVID-19 Pandemic in New York State" (Wedenoja, 2021).

For this analysis I focus on the comparing Black and Hispanic New York State residents to white residents in terms of their households, neighborhoods, education, employment, and healthcare access. These SDOH likely affected level of exposure to COVID-19 pandemic that individuals and communities experienced. Combining measures like these with knowledge of the timing and geography of COVID-19 spread and mitigation can potentially be used to causally link these and other SDOH to COVID-19 incidence and mortality.

Household

COVID-19 spreads more rapidly in crowded spaces where people interact closely (CDC, 2021) and in New York Black and Hispanic people are more likely to live in these spaces (Table 1). Statewide 57% of Black people and 62% of Hispanic people live in buildings with more than two apartments, which increases the likelihood of interacting with other households, compared to only 25% of white New Yorkers. These households are also more likely to be located in US Census Bureau defined central cities with more crowded businesses and public transportation. Only 30% of white New Yorkers live in central cities compared to 76% of Black and 73% of Hispanic New Yorkers.

[Table 1 about Here]

Black and Hispanic New Yorkers live in larger households than white New Yorkers, are more likely to live with a school aged child, and also more likely to live in a multigenerational home. Fewer than 5% of white New Yorkers live in a multigenerational home compared to 14% of Black and 12% of Hispanic New Yorkers.

Neighborhood

Black and Hispanic New Yorkers are not only more likely to live in multi-unit buildings and center cities, they live in census tracts with higher population density, more crowded housing (defined as more than one person per room), and more group housing residents per capita (Table 2). Both the dense population and crowded housing in New York City is believed to have contributed to COVID-19's rapid spread in the initial stages of the pandemic (Gonzalez-Reiche et al., 2020).

Black New Yorkers live in neighborhoods that are 2.4 times as dense and Hispanic New Yorkers live in neighborhoods that are 3.0 times as dense as those of white New Yorkers. These patterns are not solely driven by the fact that more Black and Hispanic New Yorkers live in New York City (NYC). Outside NYC Black New Yorkers live in neighborhoods 1.5 times as dense and Hispanic New Yorkers live in Neighborhoods 1.7 times as dense as white New Yorkers. The housing within neighborhood is also more likely to be classified as crowded. Black New Yorkers live in neighborhoods with twice as much crowded housing compared to white New Yorkers and Hispanic New Yorkers live in neighborhoods with three times as much crowded housing.

[Table 2 about Here]

Finally, Black New Yorkers are more likely to live in neighborhoods that have other COVID-19 risk factors including larger group quarters populations like nursing homes (Barnett and Grabowski, 2020), prisons (Kinner et al., 2020), and college dorms (Lu et al., 2020). These residential facilities have seen rapid spread throughout the pandemic. Black New Yorkers live in neighborhoods with 10% more group quarters residents per 1,000 people compared to white New Yorkers although Hispanic residents live in neighborhoods with a lower group quarters population. However, outside NYC Black and Hispanic New Yorkers live in neighborhoods with more than 50% more group quarters residents compared to white New Yorkers.

Education

There is substantial evidence that education is correlated with health but the direct causal pathways are difficult to measure. New Yorkers have higher than average education levels but there is substantial heterogeneity in educational attainment across racial and ethnic groups (Table 3). The vast majority (88%) of New Yorkers age 25 and older hold at least a high school diploma and a substantial minority (38%) hold at least a Bachelor's degree. At the household level, 94% of New Yorkers live in a household where someone has a high school diploma and 49% live in a household where someone has a bachelor's degree.

[Table 3 about Here]

These apparent high levels of educational attainment obscure significant racial and ethnic disparities in educational attainment. Only 82% of Black and 69% of Hispanic New Yorkers age 25 and older have a high school diploma compared to 94% of white New Yorkers. The difference for bachelor's degrees attainment is even greater, 43% of white New Yorkers have a

bachelor's or above compared to only 25% of Black and 21% of Hispanic New Yorkers. Only about one third of Black (35%) and Hispanic (31%) New Yorkers live in a household where at least one person has a Bachelor's degree compared to over half (56%) or white New Yorkers.

Job Characteristics

There is evidence that one of the reasons that Black and Hispanic New Yorkers were hit harder by the COVID-19 pandemic is that they are more likely to work in essential industries and jobs that are high risk for infection. (McNicholas and Poydock, 2020; Dubay et al., 2020). I identified two COVID-19 relevant job characteristics using the O*Net database: how physically close workers are to each other (proximity to others) and how often they are exposed to disease or infections. Black and Hispanic New Yorkers are significantly more likely to work in occupations where they are in close proximity to others and are more exposed to disease compared to white New Yorkers (Table 4).

[Table 4 about Here]

Black New Yorkers in particular are more likely to have "high risk occupations" defined as a job that requires working within arm's reach of another person and exposed to disease or infection at least once per month. More than 1 in 5 (23%) of Black workers were in a high risk occupation compared to 13% of white workers and 16% of Hispanic workers.[1] Black and Hispanic New Yorkers were also more likely to live in a household with a high risk worker compared to white New Yorkers. Nearly one third (29%) of Black New Yorkers and nearly a quarter (23%) of Hispanic residents lived with at least one high risk worker compared to only about one fifth (18%) of white residents.

Healthcare Access

Historically Black and Hispanic families have had less access to healthcare than white families (James et al., 2007). Over the past 10 years, even with the enactment of the Affordable Care Act which expanded health insurance access and Medicaid coverage to uncovered and under-covered families, Black and Hispanic New Yorkers have been less likely to have health insurance compared to white New Yorkers (Table 5). Nearly 95% of white New Yorkers have health insurance compared to only 91% of Black New Yorkers and 88% of Hispanic New Yorkers.

[Table 5 about Here]

Poverty is also heavily related to healthcare access generally (Ahmed et al., 2001; Litaker et al., 2005) and COVID-19 care specifically (Patel et al., 2020). Black New Yorkers are more than twice as likely to live below the poverty line compared to white New Yorkers and Hispanic New Yorkers are just over two and a half times more likely to live below the poverty line.

---

[1] The difference for white and Hispanic workers was not statistically significantly different at the 10% level.

Statistical and Econometric Methodologies for Causal Inference

Causal inference models for observational data have become increasingly common in public health research as part of the "credibility revolution" (Angrist and Pischke, 2010; Imbens and Wooldridge, 2009). Versions of these methods have been used to identify the causal effects of SDOH on health outcomes and health disparities. Jeffries et al (2019) discusses in detail many of these models and their applications with an eye toward SDOH and health disparities research.

Identifying causality without an randomized control trial (RCT) requires an "identification strategy" which is a combination of statistical tools, data, and a set of explicit reasoning as to why the correlational coefficients from these models should be interpreted as causal. These identifying assumptions generally fall into two categories: selection-on-observables (SOB) and "natural experiment." There are instances in which SDOH RCTs are possible, especially experiments focused on access to health insurance (Finkelstein et al., 2012), but they remain unusual.

SOB models, also called conditional independence assumption models, rely on the assumption that the level of "treatment" a person receives, such as the level of education, is conditionally independent, in layman's terms as good as random, when controlling for a set of that person's observable characteristics. SOB models include multiple regression with control variables, matching (Stuart, 2010), propensity score weighting (Guo and Fraser, 2014), and synthetic control methods (Abadie et al., 2010). While these methods do provide important insight, the complicated relationship between the SDOH and health disparities makes it unlikely that all relevant control variables can be captured to satisfy the conditional independence assumption even with advanced techniques.

Identification strategies that rely on natural experiments require assumptions similar to the conditional independence assumption and each model requires a slightly different version of that assumption. In the case of a "true" natural experiment the assumptions for causal interpretation are the same as an RCT. Some individuals randomly receive the "treatment" and the others do not. Because assignment was random, on average both groups had the same potential outcome under both treatment and control so directly comparing the two groups is the causal effect of the treatment. The impact of environmental factors and built environment on health disparities commonly employ natural experiments and variation including the effect of highway pollution on asthma (Bowatte et al., 2015), and the exposure to lead contaminated water in Flint, MI (Zahran et al., 2017).

Natural experimental variation in a factor of interest is generally rare. The three statistical methodologies below are appropriate for more broadly defined natural experiments and each has a different set of assumptions that must be satisfied for causal identification. This discussion is not intended as a research manual or an exhaustive catalog of these methods and

their assumptions[2], rather a primer on their potential relevance to the study of SDOH and health disparities in the context of COVID-19.

Panel Data Methods

Panel data, also called longitudinal data, includes repeated observations of the same set of individuals over multiple time periods. Long panels have relatively more temporal observations than unique individuals and wide panels have relatively more unique individuals. Panel data can be at the individual person, household, group, or geographic level and is likely to be the most important source of data to study the link between SDOH and health disparities related to COVID-19.

There are many existing panel data sources that include SDOH. Surveys include the Panel Study of Income Dynamics (PSID), the National Longitudinal Studies (NLS), and the Medical Expenditure Panel Survey (MEPS). Administrative data such as medical insurance, employment, and educational records can be adapted into panel data. Additionally, panel data for aggregate groups, for example census tracts, can be created by aggregating repeated cross sectional survey data such as the American Community Surveys (ACS), Decennial Census, American Housing Survey (AHS), Behavioral Risk Factor Surveillance System (BRFSS), and the Current Population Survey (CPS).

Repeated observations of the same individuals or groups allows for comparisons between one individual to another at a fixed point in time, and the same individual to themselves at different points in time. Fixed effects (FE) estimation compares individuals to themselves at times when they are "treated" and times when they are not. The underlying assumption is that there is an unobserved but fixed characteristic of the individual that determines whether or how much the individual is treated. As long as the characteristic is fixed, any variation in treatment overtime is "as good as random."

In the context of COVID-19, FE methods have been applied to the effect of stay at home orders on COVID-19 transmission using a panel of US counties and the timing of stay at home orders (Fowler et al., 2020) and there are substantial opportunities for additional research using similar methodology. The individual nature of FE, however, makes it slightly less attractive in understanding SDOH and health disparities because "treatment" is at the group level. Additionally, the SDOH of interest are generally persistent and it is difficult to assume that changes in an individual's SDOH overtime is only driven by an unobserved fixed factor.

Differences-in-differences (DD) estimation is generally more appropriate for the study of SDOH causality. DD models have already been used to examine the causal effect of COVID-19 on excess deaths in Italy (Ciminelli and Garcia-Mandicó, 2020), the causal effect of emergency sick leave on COVID-19 cases (Pichler et al., 2020), and the effect of COVID-19 on gender disparities

---

[2] For a practical guide see Angrist and Pischke (2008), Wooldridge (2010), Cameron and Trivedi (2005), as well as the specific following citations for each method.

in academic publications (Deryugina, 2021). DD models compare the change in outcomes for one group, which experiences a treatment over a period of time, to the change in outcomes of another group over the same period of time that did not experience the treatment. Because this model compares changes to each other, the underlying assumption for causal identification is that the two groups should have the same outcome trends pretreatment but not necessarily levels (Bertrand et al., 2004). A model with similar assumptions, but that allows for estimation across multiple groups with multiple different treatment timings, as well as time-varying effects is the event study.

DD and event studies are used extensively in SDOH related research including the causal effect of increased healthcare access (Bailey and Goodman-Bacon, 2014; Lovenheim et al., 2016). In the study of SDOH and health disparities an expansion to a triple difference model may also be valuable. Triple difference (DDD) models employ the same general strategy as DD with a third group. In the example of COVID-19 and health disparities, the third difference to compare could be racial or ethnic groups within the treatment and control groups to examine differential treatment effects.

Instrumental Variables

Instrumental variables (IV) models were originally developed to solve the simultaneous equations problem in econometrics. Economists only observe the equilibria of two equations that move simultaneously and thus cannot identify each equation individually. A partial solution to this problem was to identify a variable that should only affect either supply or demand but not both. This variable is the instrument or instrumental variable. The key assumption for causal inference is that the instrument must meet the "exclusion restriction" and only affect the outcome of interest through the instrument's effect on the independent variable. If this assumption is met, assignment into the treatment and control group is random (Angrist et al., 1996; Angrist and Krueger, 2001). IV estimation is generally done with two regressions, the first accounts for the instrument's effect on the independent variable and the second for the instrumented independent variable's effect on the outcome of interest. The most common method is two step least squares (2SLS).

In epidemiology IV models are often used to study public health interventions that do not have perfect compliance (Greenland, 2000; Hernan and Robins, 2006). In this case the instrument is the random assignment into the treatment condition and the independent variable is whether the subject complies with the treatment. Similarly, IV methods are applied to natural experiments in cases in which the exogenous natural variation does not directly affect outcomes but does affect an independent variable of interest. In the SDOH context random assignment of students to dorms, classes, and cohorts affects the peers that students interact with and those peers may affect health outcomes. IV studies have found peers affect both smoking and alcohol related health behaviors (Fletcher, 2012; Fletcher, 2010).

Identification strategies using IV require strong identification assumptions about the relationship between the instrument and the explanatory variable and the instrument and the

outcome variable and many of these assumptions are untestable. IV estimation is also complicated in cases with heterogeneous treatment effects or differential levels of compliance (Angrist and Imbens, 1995; Heckman and Vytlacil, 1999). Identification strategies using IV require substantial robustness checks and statistical validation but also qualitative subject matter knowledge into the exclusion restriction assumption, an assumption that is difficult if not impossible to test. External validity is also a concern because it is crucial to think through who the marginal case affected by the instrument is, and whether results from that marginal case are generalizable.

In the context of COVID-19 and health disparities, IV estimates are likely to be valuable in assessing the causal effect of the natural experiment that is COVID-19. Local case positivity rates, initial introduction of COVID-19 to a geography, lock down orders, and vaccine availability are all potential instruments for COVID-19 positivity or COVID-19 health behaviors. Not everyone who lived in an area with high transmission became infected just like not everyone who was subject to stay-at-home restrictions followed them. Traditional instruments that affect upstream SDOH will also continue to be valuable in the study of COVID-19. Instruments that affect where people live, their peer groups, and the type of jobs they have but do not contribute to COVID-19 transmission or mortality in other ways are valuable to the study of health disparities. IV approaches have recently been used to estimate the effect of reduced workplace presence on COVID-19 deaths (McLaren and Wang, 2020) and the effect of political institutions on initial COVID-19 deaths (Cepaluni, 2021).

Regression Discontinuity

Regression discontinuity (RD) designs exploit variance in assignment to treatment and control groups based on clear thresholds that subjects are unable to fully manipulate relying on the assumption that for the group of individuals just around the threshold, whether they fall just above or just below is as good as random (Thistlethwaite and Campbell, 1960). In a sharp RD design the threshold is fully determinant of treatment, all individuals who fall just above are treated and all of those who fall just below are not (Imbens and Lemieux, 2008). In a fuzzy RD design, the threshold is not determinant, rather the probability of treatment changes dramatically at the threshold so the threshold becomes an instrumental variable for treatment status. Fuzzy RD designs are much more common.

RD designs have many potential applications in public health and healthcare research but are generally underutilized (Moscoe et al., 2015; Venkataramani et al., 2016). The discontinuity in access to a parent's health insurance at age 26 that resulted from the affordable care act has been used to estimate the effect of health insurance on health behaviors, specifically seeking preventative medical and emergency care (Anderson et al., 2012). The effect of retirement on health behaviors has similarly employed an age discontinuity (Muller and Shaikh, 2018).

Age related RD designs will likely be appropriate for COVID-19 research as well. Most states rolled out vaccination availability by age group which presents a perfect structure for age-based fuzzy RD design for the effect of vaccines on COVID-19 incidence and mortality as well as other

health behaviors including attending preventative screenings or dental appointments. The rules and cutoffs of the three COVID-19 stimulus payments, the paycheck protection program (PPP) loans and grants, and the advanced child tax credit payments are also potential discontinuities. These present the opportunity not only to study the effect of income on COVID-19 specific health outcomes, but also the effect of income on longer term and non-COVID health outcomes.

The Need for High Quality Data

The methods described above can only be successful in identifying the causal relationship between the SDOH and health disparities in the presence of high quality data. Many existing sources of data can be expanded, adapted, or linked in order to successfully use these strategies. Below are four strategies for building datasets that allow for causal identification of SDOH measures on health disparities in the context of COVID-19.

Measurable and Moveable SDOH measures

While SDOH are broad and multifaceted, in order to address their impact on health disparities through causal inference methods they must have accurate and specific measures in datasets. Healthy People 2030 includes a list of these measures and targets for their improvement. Further, these measures must be moveable through policy. Some health disparities, including the prevalence of sickle-cell disease, are the result of inherited genes and are thus not moveable and should not be considered for this type of research. However, policy can be used to influence the characteristics of where people live, work, and go to school. The high school graduation rate, unemployment rate, average wage, and concentration of crowded housing are all sensitive to policy interventions.

Adding COVID-19 questions to ongoing surveys

As discussed above in the panel data methodology section, there are a significant number of panel and repeated cross-sectional surveys in the US that are employed to study SDOH and health disparities. Adding questions about COVID-19 to these surveys will be invaluable in studying the long term effects of the pandemic on health disparities and the role of SDOH. Topics could include COVID test positivity, job loss due to COVID-19, stimulus payments, remote learning and work from home. These can be outcomes or leveraged with an IV model to understand the long term effects of COVID-19 on future health disparities.

A COVID-19 prevalence and policy database

The three methods of causal inference above leverage some form of natural experiment. In order for them to be valid or testable that natural variation must be measured accurately and specifically for any groups of interest. While COVID-19 case rates are not fully exogenous to a location or its policies, there is a good argument to be made that the date of initial introduction of COVID-19 to an area is exogenous or that, controlling for certain variables in an IV or RD

design, the case rate experienced by an individual is. The timing of cases is of particular importance to FE and DD research designs.

The timing of COVID mitigation efforts, such as non-essential business closures, stay at home orders, mask mandates, and rules about indoor dining, are also potential exogenous sources of variation. Special attention should be paid to cataloging restrictions that are "rules based" including staggered re-openings determined by case rate bench marks or mandates that were invalidated by external forces like state or court intervention.

Finally, any COVID-19 policy that was based on a specific eligibility cutoff should be documented with all cutoffs and the rules (and exceptions to those rules) included. The obvious example is the age-based roll out of the COVID-19 vaccine and the peculiarity that eligible age groups differed by state on any given date. Additional COVID-19 policies with cutoffs should not be overlooked including stimulus payments, unemployment benefits, child tax credit payments, rental assistance, and PPP loans. Access to these programs my alter SDOH and have long term health impacts.

Leveraging existing data sources in new ways

The effect of COVID-19, especially in the first wave, was profoundly localized. Throughout the pandemic case rates, hospital occupancy rates, and mortality rates fluctuated wildly over very small geographies. Using the methods above these geographic variations and SDOH in these geographies can be leveraged to study both COVID-19 and other health disparities. This necessitates building small-geography panel data that incorporates SDOH, COVID-19 measures, and health behaviors and outcomes. This can include aggregating multiple survey and administrative datasets over the same geography and combining them with COVID-19 related data. The pre-COVID SDOH measures discussed above do just that for the large geography of New York. These measures were pulled from five different national surveys. Researchers have already started to link small geography SDOH to COVID-19 health outcomes in New York City (Wadhera et al., 2020), Chicago (Kim and Bostwick, 2020), and Louisiana (Price-Haywood et al., 2020).

Linking administrative data also presents an important opportunity to study SDOH in the context of COVID-19. The pandemic profoundly altered the way that people live and there are likely to be long term consequences to health disparities from the economic and social shifts that the pandemic incited. Administrative data on employment changes, K-12 school enrollment, college enrollment, income, and healthcare access can be used to examine other long-term health disparities. COVID-19 mortality and morbidity is an important outcome but the structural changes to SDOH driven by the pandemic will likely have far reaching consequences for health disparities.

Conclusion

The earliest research into disparities in COVID-19 impact has focused almost exclusively on COVID-19 as an outcome and most studies have been speculative about the relationship between disparities in the social determinants of health and COVID-19 disparities. Measuring the causal effect of SDOH on health disparities in any context is extremely difficult because the relationship is complex and overlapping.

The COVID-19 pandemic and its effect on public policy has the potential to provide new causal evidence on the relationship between SDOH and health disparities in the US. Pre-pandemic data is suggestive that SDOH were the primary reason why Black and Hispanic New Yorkers were disproportionately affected by COVID-19 but rigorous causal inference statistical methods are needed to fully understand the relationship between SDOH and COVID-19 disparities.

The COVID-19 pandemic has generated a great deal of natural experiments that not only affect COVID-19 transmission and mortality but also education, income, housing, and employment that are likely to have longer term effects on health disparities. In order to study both these long and short term effects we need high quality sources of health and SDOH data. This may include adding new questions to existing surveys, cataloging COVID-19 policies and rules, and linking administrative datasets.

Works Cited

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105, no. 490 (2010): 493-505.

Abrams, Elissa M., and Stanley J. Szefler. "COVID-19 and the impact of social determinants of health." *The Lancet Respiratory Medicine* 8, no. 7 (2020): 659-661.

Ahmed, Syed M., Jeanne P. Lemkau, Nichol Nealeigh, and Barbara Mann. "Barriers to healthcare access in a non-elderly urban poor American population." *Health & social care in the community* 9, no. 6 (2001): 445-453.

Alvidrez, Jennifer, Dorothy Castille, Maryline Laude-Sharp, Adelaida Rosario, and Derrick Tabor. "The national institute on minority health and health disparities research framework." *American Journal of Public Health* 109, no. S1 (2019): S16-S20.

Anderson, Michael, Carlos Dobkin, and Tal Gross. "The effect of health insurance coverage on the use of medical services." *American Economic Journal: Economic Policy* 4, no. 1 (2012): 1-27.

Angrist, Joshua D., and Alan B. Krueger. "Instrumental variables and the search for identification: From supply and demand to natural experiments." *Journal of Economic perspectives* 15, no. 4 (2001): 69-85

Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly harmless econometrics*. Princeton university press, 2008.

Angrist, Joshua D., and Jörn-Steffen Pischke. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of economic perspectives* 24, no. 2 (2010): 3-30.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91, no. 434 (1996): 444-455.

Angrist, Joshua, and Guido Imbens. "Identification and estimation of local average treatment effects." (1995).

Artiga, Samantha, Latoya Hill, Sweta Hadler. "COVID-19 Cases and Deaths by Race/Ethnicity: Current Data and Changes Over Time" Kaiser Family Foundation Issue Brief published October 8, 2021, https://www.kff.org/racial-equity-and-health-policy/issue-brief/covid-19-cases-and-deaths-by-race-ethnicity-current-data-and-changes-over-time/

Bailey, Martha J., and Andrew Goodman-Bacon. "The War on Poverty's experiment in public medicine: Community health centers and the mortality of older Americans." *American Economic Review* 105, no. 3 (2015): 1067-1104.

Barnett, Michael L., and David C. Grabowski. "Nursing homes are ground zero for COVID-19 pandemic." In *JAMA Health Forum*, vol. 1, no. 3, pp. e200369-e200369. American Medical Association, 2020.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?." *The Quarterly journal of economics* 119, no. 1 (2004): 249-275.

Bowatte, Gayan, C. Lodge, Adrian J. Lowe, Bircan Erbas, Jennifer Perret, Michael J. Abramson, M. Matheson, and Shyamali C. Dharmage. "The influence of childhood traffic-related air pollution exposure on asthma, allergy and sensitization: a systematic review and a meta-analysis of birth cohort studies." *Allergy* 70, no. 3 (2015): 245-256.

Braveman, Paula, Susan Egerter, and David R. Williams. "The social determinants of health: coming of age." *Annual review of public health* 32 (2011): 381-398.

Burström, Bo, and Wenjing Tao. "Social determinants of health and inequalities in COVID-19." (2020): 617-618.

Cameron, A. Colin, and Pravin K. Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

Cepaluni, Gabriel, Michael T. Dorsch, and Réka Branyiczki. "Political regimes and deaths in the early stages of the COVID-19 pandemic." *Journal of Public Finance and Public Choice* (2021).

Ciminelli, Gabriele, and Sílvia Garcia-Mandicó. "COVID-19 in Italy: an analysis of death registry data." *Journal of Public Health* 42, no. 4 (2020): 723-730.

Deryugina, Tatyana, Olga Shurchkov, and Jenna Stearns. "Covid-19 disruptions disproportionately affect female academics." In *AEA Papers and Proceedings*, vol. 111, pp. 164-68. 2021.

"Disparities in Deaths from COVID-19," Centers for Disease Control and Prevention, last modified December 10, 2020, https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-deaths.html

Dubay, Lisa, Joshua Aarons, K. Steven Brown, and Genevieve M. Kenney. "How risk of exposure to the coronavirus at work varies by race and ethnicity and how to protect the health and well-being of workers and their families." *Health Policy Center, Research Report, December* (2020).

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. "The Oregon health insurance experiment: evidence from the first year." *The Quarterly journal of economics* 127, no. 3 (2012): 1057-1106.

Fletcher, Jason M. "Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach." *Journal of Population Economics* 25, no. 4 (2012): 1265-1286.

Fletcher, Jason M. "Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects." *Health economics* 19, no. 4 (2010): 466-484.

Gonzalez-Reiche, Ana S., Matthew M. Hernandez, Mitchell J. Sullivan, Brianne Ciferri, Hala Alshammary, Ajay Obla, Shelcie Fabre et al. "Introductions and early spread of SARS-CoV-2 in the New York City area." *Science* 369, no. 6501 (2020): 297-301.

Greenland, Sander. "An introduction to instrumental variables for epidemiologists." *International journal of epidemiology* 29, no. 4 (2000): 722-729.

Guo, Shenyang, and Mark W. Fraser. *Propensity score analysis: Statistical methods and applications*. Vol. 11. SAGE publications, 2014.

Heckman, James J., and Edward J. Vytlacil. "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the national Academy of Sciences* 96, no. 8 (1999): 4730-4734

Hernán, Miguel A., and James M. Robins. "Instruments for causal inference: an epidemiologist's dream?." *Epidemiology* (2006): 360-372.

"How COVID-19 Spreads," Centers for Disease Control and Prevention, updated July 14, 2021, https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html

Imbens, Guido W., and Jeffrey M. Wooldridge. "Recent developments in the econometrics of program evaluation." *Journal of economic literature* 47, no. 1 (2009): 5-86.

Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." *Journal of econometrics* 142, no. 2 (2008): 615-635.

Jeffries, Neal, Alan M. Zaslavsky, Ana V. Diez Roux, John W. Creswell, Richard C. Palmer, Steven E. Gregorich, James D. Reschovsky et al. "Methodological approaches to understanding causes of health disparities." *American journal of public health* 109, no. S1 (2019): S28-S33.

Kim, S. J., and W. Bostwick. "Social Vulnerability and Racial Inequality in COVID-19 Deaths in 508 Chicago." *Health education & behavior: the official publication of the Society for Public Health* 509: 509-13.

Kinner, Stuart A., Jesse T. Young, Kathryn Snow, Louise Southalan, Daniel Lopez-Acuña, Carina Ferreira-Borges, and Éamonn O'Moore. "Prisons and custodial settings are part of a comprehensive response to COVID-19." *The Lancet. Public Health* 5, no. 4 (2020): e188.

Kneipp, Shawn M., Todd A. Schwartz, Denise J. Drevdahl, Mary K. Canales, Sheila Santacroce, Hudson P. Santos Jr, and Ruth Anderson. "Trends in health disparities, health inequity, and social determinants of health research: a 17-year analysis of NINR, NCI, NHLBI, and NIMHD funding." *Nursing research* 67, no. 3 (2018): 231-241.

Koh, Howard K., Sarah C. Oppenheimer, Sarah B. Massin-Short, Karen M. Emmons, Alan C. Geller, and Kasisomayajula Viswanath. "Translating research evidence into practice to reduce health disparities: a social determinants approach." *American journal of public health* 100, no. S1 (2010): S72-S80.

Litaker, David, Siran M. Koroukian, and Thomas E. Love. "Context and healthcare access: looking beyond the individual." *Medical care* (2005): 531-540.

Lovenheim, Michael F., Randall Reback, and Leigh Wedenoja. *How does access to health care affect teen fertility and high school dropout rates? Evidence from school-based health centers*. No. w22030. National Bureau of Economic Research, 2016.

Lu, Hannah, Cortney Weintz, Joseph Pace, Dhiraj Indana, Kevin Linka, and Ellen Kuhl. "Are college campuses superspreaders? A data-driven modeling study." *Computer Methods in Biomechanics and Biomedical Engineering* 24, no. 10 (2021): 1136-1145.

McLaren, John, and Su Wang. *Effects of reduced workplace presence on COVID-19 deaths: An Instrumental-variables approach*. No. w28275. National Bureau of Economic Research, 2020.

McNicholas, Celine, and Margaret Poydock. "Who are essential workers? A comprehensive look at their wages, demographics, and unionization rates." *Economic Policy Institute* 19 (2020).

"Minority Health and Health Disparities: Definitions and Parameters," National Institutes of Health: National Institute on Minority Health and Health Disparities, updated March 20, 2021, https://www.nimhd.nih.gov/about/strategic-plan/nih-strategic-plan-definitions-and-parameters.html

Moscoe, Ellen, Jacob Bor, and Till Bärnighausen. "Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice." *Journal of clinical epidemiology* 68, no. 2 (2015): 132-143

Müller, Tobias, and Mujaheed Shaikh. "Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design." *Journal of health economics* 57 (2018): 45-59.

National Institute of Child Health, and Human Development (US). *Health disparities: Bridging the gap*. The Development, 2000.

Patel, J. A., F. B. H. Nielsen, A. A. Badiani, S. Assi, V. A. Unadkat, B. Patel, R. Ravindrane, and H. Wardle. "Poverty, inequality and COVID-19: the forgotten vulnerable." *Public health* 183 (2020): 110.

Penman-Aguilar, Ana, Makram Talih, David Huang, Ramal Moonesinghe, Karen Bouye, and Gloria Beckles. "Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity." *Journal of public health management and practice: JPHMP* 22, no. Suppl 1 (2016): S33.

Pichler, Stefan, Katherine Wen, and Nicolas R. Ziebarth. "COVID-19 Emergency Sick Leave Has Helped Flatten The Curve In The United States: Study examines the impact of emergency sick leave on the spread of COVID-19." *Health Affairs* 39, no. 12 (2020): 2197-2204.

Price-Haywood, Eboni G., Jeffrey Burton, Daniel Fort, and Leonardo Seoane. "Hospitalization and mortality among black patients and white patients with Covid-19." *New England Journal of Medicine* 382, no. 26 (2020): 2534-2543.uses hospitalization data in a predominantly black and white area to look at the progression of COVID positive patients

Schillinger, Dean. "The intersections between social determinants of health, health literacy, and health disparities." *Studies in health technology and informatics* 269 (2020): 22.

Singu, Sravani, Arpan Acharya, Kishore Challagundla, and Siddappa N. Byrareddy. "Impact of social determinants of health on the emerging COVID-19 pandemic in the United States." *Frontiers in public health* 8 (2020): 406.

Stuart, Elizabeth A. "Matching methods for causal inference: A review and a look forward." *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, no. 1 (2010): 1.

Thistlethwaite, Donald L., and Donald T. Campbell. "Regression-discontinuity analysis: An alternative to the ex post facto experiment." *Journal of Educational psychology* 51, no. 6 (1960): 309.

Turner-Musa, Jocelyn, Oluwatoyin Ajayi, and Layschel Kemp. "Examining social determinants of health, stigma, and COVID-19 disparities." In *Healthcare*, vol. 8, no. 2, p. 168. Multidisciplinary Digital Publishing Institute, 2020.

Venkataramani, Atheendar S., Jacob Bor, and Anupam B. Jena. "Regression discontinuity designs in healthcare research." *Bmj* 352 (2016).

Wadhera, Rishi K., Priya Wadhera, Prakriti Gaba, Jose F. Figueroa, Karen E. Joynt Maddox, Robert W. Yeh, and Changyu Shen. "Variation in COVID-19 hospitalizations and deaths across New York City boroughs." *Jama* 323, no. 21 (2020): 2192-2195.

Wedenoja, Leigh. "Measuring Disparities in the Social Determinants of Health in Relation to the COVID-19 Pandemic in New York State" Rockefeller Institute of Government Report published March 2021, https://rockinst.org/issue-area/measuring-health-disparities/

Wong, Mitchell D., Martin F. Shapiro, W. John Boscardin, and Susan L. Ettner. "Contribution of major diseases to disparities in mortality." *New England Journal of Medicine* 347, no. 20 (2002): 1585-1592.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

Wright, Philip G. *Tariff on animal and vegetable oils*. Macmillan Company, New York, 1928.

Zahran, Sammy, Shawn P. McElmurry, and Richard C. Sadler. "Four phases of the Flint water crisis: Evidence from blood lead levels in children." *Environmental Research* 157 (2017): 160-172.

Zimmerman, Emily, and Steven H. Woolf. "Understanding the relationship between education and health." *NAM Perspectives* (2014).

Tables

Table 1: Household

	Population	White	Black		Hispanic	
Household in Structure with >2 Apartments	38.2%	25.1%	57.1%	***	62.0%	***
Household Located in Central City	48.2%	29.7%	76.1%	***	72.5%	***
Average Number of Household Members	3.3	3.1	3.4	***	3.8	***

*** = average is statistically different from non-Hispanic white at 1% level, **=5% level, *=10% level

SOURCE: IPUMS-CPS, University of Minnesota, www.ipums.org.

Table 2: Neighborhood

	Population	White	Black	Hispanic
Average Census Tract Density Per Square Mile	29,337	17,072	40,812	50,816
Average Percent Crowded (>1 person/room) Housing Units	5.64%	3.25%	7.41%	9.91%
Group Quarters per 1000 Tract Residents	30	31	34	25

SOURCE: US Census Bureau. (2020). Planning Database with 2010 Census and 2014 – 2018 American Community Survey Data. [Data file]. Retrieved from https://www.census.gov/topics/research/guidance/planning-databases.html

Table 3: Education

	Population	White	Black		Hispanic	
25+ with HS diploma+	88%	94%	82%	***	69%	***
25+ with Bachelor's Degree+	38%	43%	25%	***	21%	***
HH Member has HS Diploma+	94%	97%	91%	***	83%	***
HH Member has Bachelor's+	49%	56%	35%	***	31%	***

*** = average is statistically different from non-Hispanic white at 1% level, **=5% level, *=10% level

IPUMS-CPS, University of Minnesota, www.ipums.org.

Table 4: Employment

	Population	White	Black		Hispanic	
Proximity To Others Score	62.5	61.1	65.7	***	64.2	***
Exposure to Disease Score	24.0	21.9	31.1	***	24.7	**
High Risk Occupation (disease>50, prox >75)	15.4%	13.0%	22.8%	***	16.0%	
Lives in Household with High Risk Worker	21.1%	17.7%	29.1%	***	23.4%	***

*** = average is statistically different from non-Hispanic white at 1% level, **=5% level, *=10% level

IPUMS-CPS, University of Minnesota, www.ipums.org.

National Center for O*NET Development. O*NET OnLine. Retrieved January 29, 2021, from https://www.onetonline.org/

Table 5: Healthcare Access

	Population	White	Black		Hispanic	
Below Poverty Line	14.5%	9.4%	22.5%	***	24.0%	***
Has Health Insurance	92.8%	94.7%	90.7%	***	88.7%	***

*** = average is statistically different from non-Hispanic white at 1% level, **=5% level, *=10% level

IPUMS-CPS, University of Minnesota, www.ipums.org.