Title: How are key demographic indicators related to COVID-19 reported in the United States: A data review protocol[1]

Date: First draft: April 26 2020; Second draft: June 29, 2020; Third draft: July 13, 2020

Authors: Charlotte Brasseux[1], Thoai D Ngo[2,3], Mingqi Song[4]

Affiliations:
[1]MPH candidate at Mailman School of Public Health at Columbia University
[2]Poverty, Gender and Youth Program, Population Council, New York, New York, USA
[3]The GIRL Center, Population Council, New York, New York, USA
[4] MPA, School of International and Public Affairs at Columbia University

Contacts:
Thoai D Ngo, PhD, MHS
Program Director
Poverty, Gender and Youth Program
Population Council
tngo@popcouncil.org.

Charlotte Brasseux, MPH candidate
Mailman School of Public Health at Columbia University
cab2315@cumc.columbia.edu

Mingqi Song, MPA
ms5695@columbia.edu

POPULATION COUNCIL
IDEAS. EVIDENCE. IMPACT.
One Dag Hammarskjold Plaza
New York, NY 10017
www.popcouncil.org

Abstract
The COVID-19 pandemic is evolving fast and affecting rich and poor countries; however, the social determinants of the health outcomes associated with COVID-19 have not been well characterized. While there is increasing discussion of COVID-19 disparities in the media, no systematic compilation of data exists that synthesizes what and how official reporting platforms disaggregate demographic indicators of COVID-19 testing, cases, hospitalizations, recoveries and deaths. This review will provide a comprehensive overview of what and how federal and local health officials report COVID-19 cases. Having this information nationally and sub-nationally can help health officials to deploy a more targeted response effort such as testing, treatment, and contact tracing. This information could be useful for future vaccine development. This data review can reveal gaps in our public health data system, which can lead to recommendations on how to improve data standardization, disaggregation, and reporting, particularly for tracking outbreaks. This protocol may be replicated in other countries to understand how demographic indicators are being reported and their relationship to COVID-19 health outcomes.

---

[1] This document is evolving due to the nature of the COVID-19 response and will be updated as needed.

POPULATION COUNCIL
Ideas. Evidence. Impact.

Background

The World Health Organization (WHO) Commission on Social Determinants of Health asserts that political, social, and economic forces shape the circumstances in which people live. These circumstances give rise to avoidable health inequities, both within and between countries, which are strongly linked to degrees of social disadvantage[1]. In an article from the *Annual Review of Public Health*, Adler and Rehkopf describe that "physical and social environments, including a person's home, school, work, neighborhood, and community, vary by socio-economic status (SES) and affect the likelihood of individuals' exposure to both health-damaging conditions and health-protecting resources"[2]. Factors including race/ethnicity, gender, income, occupation, education level, and geographic location all have an impact on health outcomes, such as morbidity and mortality rates. The first study in the United States (US) to examine socioeconomic disparities within race/ethnicity using individual-level data from the 1960 matched records of people 25 years of age and older. This study found that compared to whites, age-adjusted, all-cause mortality rates were 34% higher for nonwhite females and 20% higher for males. The same study found that males and females who are the least educated experience 64% and 105% higher mortality rates, respectively, compared to those most educated[2]. Similar trends are apparent in maternal mortality rates. Recent data from the Center for Disease Control and Prevention (CDC) shows that non-Hispanic black women have a rate of 37.1 deaths per 100,000 live births, compared to non-Hispanic white women who have a rate of 14.7 deaths per 100,000 live births[3]. Despite the recognition that addressing these disparities is critical to tackling health inequity and protecting public health, data systems, research and policies that directly tackle these key social determinants continue to lag behind in the US and globally.

During the ongoing COVID-19 pandemic, the data on COVID-19 confirmed cases and deaths in the US are increasingly revealing that people of color are more severely affected by the novel coronavirus. On April 7, the surgeon general first acknowledged the racial disparities of COVID-19 based on case reports from a few states, including Louisiana, Michigan, and Illinois[4]. Analysis of these reports suggested that majority-black counties had three times the infection rate and six times the death rate as majority-white counties[4]. As more states began to release socio-demographic data related to COVID-19, similar trends appeared. In New York, black people make up nine percent of the population but account for 17 percent of the deaths[5]. In addition to the racial disparities, COVID-19 is affecting poor people more severely and is infecting and killing more men than women, is killing older populations compared to young age groups, and is infecting urban populations more than rural ones[6, 7]. On April 17 2020, the CDC released the first report on COVID-19 containing socio-demographic data collected from only 14 states during the month of March. The report stated that only 18% of residents in the sample population were black, yet blacks made up 33% of those hospitalized with COVID-19. In comparison, while 59% of residents in the sample population were white, they made up just 45% of those hospitalized with COVID-19[8]. Public health and medical experts are stressing the fact that people of color, in general, are not more susceptible to the virus, but they are more susceptible to experiencing more severe illness and worse health outcomes as a result of the virus due to the existing health inequities in the US[9].

Historically, epidemics disproportionately affect vulnerable communities, including those living in poor and crowded neighborhoods, women and girls, gender minority individuals, and people

living in poverty. In the article *Health Inequities and Infectious Disease Epidemics*, Quinn and Kumar describe, “poverty, inequality, and social determinants of health create conditions for the transmission of infectious diseases, and existing health disparities or inequalities can further contribute to unequal burdens of morbidity and mortality”[10]. During a pandemic, these inequities become more visible due to the greater impact on vulnerable populations. During the H1N1 pandemic, Quinn and Kumar found that those who were at the highest risk of exposure to H1N1 were those of lower socioeconomic status and were likely to have less access to care if they contracted the disease[11]. Similarly, Dr. Anthony Fauci compared COVID-19 to the HIV pandemic, where HIV affected the gay communities at higher rates compared to the general population, and the COVID-19 pandemic is bringing attention to the racial disparities on health in the US[9]. 2018 data from the CDC continues to show that Black/African American gay, bisexual, and other men were more affected by HIV than any other group in the US[12].

Understanding the socio-demographic trends of COVID-19 could help with deploying preventive and mitigation measures as well as treatment and relief. Having information on who is most affected by this pandemic will help state level governments and local officials to plan a response and offer more support to communities and individuals who need it the most. Further, extensive research has demonstrated that factors including sex, pregnancy, and comorbid conditions can impact the immune response to vaccines[13]. Therefore, understanding socio-demographic factors of those most affected by COVID-19 is critical in the development of an effective vaccine. More generally, Dr. Lisa A. Cooper, Professor at the Johns Hopkins Bloomberg School of Public Health and Johns Hopkins School of Medicine, explains that “health and public health professionals, administrators, employers, policymakers, and even community advocates can use these data to determine how best to use the vast resources we have in this country to improve the lives of our people”[14].

Prevention, mitigation, treatment and relief efforts cannot occur without adequate data and reporting of socio-demographic factors of COVID-19 cases. As a result of the early data suggesting extreme racial disparities in the COVID-19 pandemic, the federal government is being pressured to not only release racial data on cases and deaths, but also release a plan to “blunt the devastation” on people and communities of color[15]. Although, the WHO does not produce official analyses and routine reporting on the socio-demographic factors related to COVID-19 confirmed cases, hospitalizations, deaths and recoveries. The CDC only began reporting this information in 28% of states on April 17, almost three months after the first reported case in the US. In the US, data revealing racial disparities of COVID-19 come from state governments reports; however, these reports are not standardized between states, and the available information is not systematically synthesized at the national level. Together, this highlights the failure of the US public health data system to capture and analyze this information in order to contribute to the test, treat, trace, and isolate initiative, which is needed to reduce and suppress the spread of coronavirus. Capturing and analyzing this information is critical in recording and tackling health inequity in the United States.

Objectives

We are conducting a data review of COVID-19 to examine how cases are reported in the US and how this reporting has changed over time with the following objectives:

1. To assess whether official or reliable case reports of COVID-19 are disaggregated by key demographic indicators, age, sex, race/ethnicity, education level, geography, economic status, and underlying conditions.

2. To assess how COVID-19 affects different population groups by cases, testing, hospitalizations, recoveries and deaths.

3. To describe how the patterns of case reports of COVID-19 change over time and from earlier epicenters to recent epicenters since the first confirmed case.

Methods

This data review examined COVID-19 reports in the US using 70 data sources from the Centers for Disease Control and Prevention (CDC) and health departments across 50 states, US territories, and ten major cities. The data review protocol is published on Dataverse[15].

Data Sources Included

We included data from 70 sources in our review (the CDC, and health departments of 50 states, 10 cities, and 9 jurisdictions). We used data from each source's health department website because we were specifically looking for how national, state, and local health departments report on COVID-19. For each source, information was extracted from three different types of data reports: primary, secondary, and tertiary data. Primary reports are daily or weekly situation reports released by local, state, and the national government. Secondary reports are reports available on each health department's COVID-19-specific website or COVID-19 dashboard. Tertiary reports are reports presented by a third-party entity linked to each health department's website.

We did not include reports from news articles, published papers (e.g., papers synthesizing and analyzing governments reports, or information from tracking sites that compile government data (e.g., Johns Hopkins University Coronavirus Resource Center COVID-19 Case Tracker). We excluded reports from county health departments.

Search Strategy

The CDC webpage reporting COVID-19 cases in the US provides links to each US jurisdiction's health department, from which the CDC retrieves its national data[16]. Using this resource, we accessed each jurisdiction's health department webpage to determine what and how each health department reported on COVID-19. The US jurisdictions identified through the CDC include 50 states and 9 territories. The territories are American Samoa, Guam, Marshall Islands, Micronesia, Northern Marianas, Palau, Puerto Rico, Virgin Islands, and the District of Columbia. We also used the Johns Hopkins University Coronavirus Resource Center website to identify the ten major cities with the highest number of confirmed COVID-19 cases (between May 30, 2020)[1]. The ten cities identified were New York City, Los Angeles, Philadelphia, Miami, New Orleans, Chicago, Detroit, Seattle, Boston, and Houston. We then searched Google for each city's health department webpage to determine what and how each city health department reported on COVID-19.

POPULATION COUNCIL
Ideas. Evidence. Impact.

In this review, 'sources' refer to the health department webpage for the CDC, 50 states, 9 territories, and 10 cities, for a total of 70 sources. For each source, we looked reporting on five key outcomes on interest and disaggregation of these outcomes by seven key demographic indicators, for a total of 40 indicators and outcomes. We also examined which sources included intersectional analysis in their reports by looking for disaggregation of outcomes by the intersection of more than one indicator (i.e. cases disaggregated by age, broken down by sex. We recorded which sources included any intersectional analysis for at least one outcome. To review each source, we first looked for a published daily or weekly COVID-19 report. We then looked for information presented separately from an official report, including information on the health department's webpage and/or dashboard designated to COVID-19. Finally, we looked at whether the health department provided links to external sources with additional data on COVID-19 for that city, state, or territory.

We began the review of the sources on May 14, 2020 and completed the review on May 30, 2020. One researcher recorded how many and which of the 40 outcomes and indicators each source reported, and if each source reported intersectionality. Another researcher verified this information. Any disagreements were discussed among all three researchers.

Outcomes of interest

We extracted data on four COVID-19 outcomes (cases, hospitalizations, recoveries, and deaths) and testing. We refer to these as the five key outcomes.

Testing refers to the total amount of positive and negative test results from state, commercial, and clinical laboratory tests. Tests include Polymerase Chain Reaction (PCR) and serology tests for SARS-CoV-2. Because someone can get tested more than once, the number of tests does not refer to the number of people who have been tested.

Cases refer to confirmed and probable COVID-19 cases. According to the CDC, a confirmed case is defined by meeting confirmatory laboratory evidence for COVID-19. A probable case is defined by i) meeting clinical criteria AND epidemiologic evidence with no confirmatory laboratory testing performed for COVID-19; or ii) meeting presumptive laboratory evidence AND either clinical criteria OR epidemiologic evidence; or iii) meeting vital records criteria with no confirmatory laboratory testing performed for COVID19[16].

Hospitalizations refer to reported hospital beds and/or intensive care units (ICU) occupied by confirmed or probable COVID-19 cases.

Recoveries refer to those who have presumably recovered from a confirmed COVID-19 diagnosis. The definition of recoveries varies between sources. The different definitions of recoveries are described in Table 1.

Deaths refer to the confirmed and probable deaths due to COVID-19. Based on the CDC definition, a confirmed death is a death caused by COVID-19 as confirmed by laboratory evidence. A probable death is defined by i) meeting clinical criteria AND epidemiologic evidence with no confirmatory laboratory testing performed for COVID-19; or ii) meeting presumptive laboratory evidence AND either clinical criteria OR epidemiologic evidence; or iii) meeting vital records criteria with no confirmatory laboratory testing performed for COVID19[16].

POPULATION COUNCIL
Ideas. Evidence. Impact.

Demographic indicators and underlying conditions

We consider demographic indicators as age, sex, race/ethnicity, education level, geography, economic status, and underlying conditions. In this data review, we examine the disaggregation of these indicators by the five outcomes. Geography is defined as the county (or zip code for cities) in which the case was reported. This information allows for understanding the distribution of the outcomes in different poor versus less-poor counties.

The CDC reports the most common underlying conditions among COVID-19 cases in the US as cardiovascular disease, diabetes, and chronic lung disease[17]. The CDC also reports the following as underlying conditions: asthma, autoimmune disease, gastrointestinal/liver disease, hypertension, immune suppression, metabolic disease, neurologic disease, obesity, pregnancy, renal disease, and other diseases[18].

Analysis

We created an Excel database to record whether each source reports on the five outcomes and the seven demographic indicators for each outcome. Each source is given an 'overall completeness score", ranging from 0 to 40, representing how many of the 40 indicators and outcomes each source reported. A total score of 40 would indicate that the source is "complete", with "complete" defined as: all indicators and outcomes are incorporated into the data report.

We also assessed how many of the 7 indicators were reported to generate a score representing the completeness of each outcome. For each data source, each outcome is given an "outcome completeness score" ranging from 0 to 7, representing how many of the 7 indicators are reported for each outcome.

For each data source, we also assessed whether the demographic indicators were cross tabulated by outcomes (i.e. cases by sex broken down by age), which is how we assessed intersectionality. We considered a source to report on intersectionality if they cross-tabulated at least two indicators by at least one outcome. We also recorded what indicators were cross tabulated by what outcomes for each source.

We also conducted a sub-analysis, which excluded education level, economic status, and underlying conditions as indicators, resulting in a total of 25 indicators and outcomes instead of 40 indicators and outcomes. In this sub-analysis, the "overall completeness score" ranged from 0 to 25 instead of 0 to 40, and the "outcome completeness score" ranged from 0 to 4 instead of 0 to 7.

Implications of our review

This review will provide a comprehensive overview of what and how federal and local health officials report COVID-19 cases. Having this information nationally and sub-nationally can help health officials to deploy a more targeted response effort and can help reveal information that could be useful for testing, treatment, and vaccine development efforts. Further, this data review may reveal gaps in our public health data system in which we can make recommendations on how to improve data standardization, disaggregation, and reporting,

particularly for tracking outbreaks. This protocol may be replicated to other countries to understand how demographic indicators are being reported and their relationship to COVID-19 cases.

Proposed Tables:

1. Description of data sources included in the review
2. Description of overall completeness of reporting by national, state, city, and territory health departments
3. Description of how national, state, city, and territory health departments report COVID-19 outcomes (testing, cases, hospitalizations, recoveries, deaths) are reported by demographic indicators
4. An analysis of disparities of COVID-19 cases nationally and by states
5. An analysis of how reporting of socio-demographic characteristics changes over the course of the COVID-19 epidemic

POPULATION COUNCIL
Ideas. Evidence. Impact.

Table 1. Definitions of recoveries by source

Source	Definition
Alabama, Alaska, Arkansas, Delaware, Hawaii, Iowa, Kentucky, Maryland, New Hampshire, New Mexico, Washington, West Virginia, New York City, Los Angeles, Philadelphia, Boston, Houston	Not specified
Mississippi, Louisiana	1) it has been 14 days or more since the case tested positive, if they were not hospitalized; 2) it has been 21 days or more since the case tested positive, if they were hospitalized or hospitalization was unknown
Idaho, Michigan	Persons with a confirmed COVID-19 diagnosis who are alive 30 days post-onset (or referral date if onset is not available).
Ohio	Cases with a symptom onset date >21 days prior who are not deceased
Maine, South Dakota	Met the released from isolation requirements defined by CDC. The requirements are: at least 3 days (72 hours) have passed since recovery, defined as resolution of fever without the use of fever-reducing medications and improvement in respiratory symptoms (e.g., cough, shortness of breath); and, at least 10 days have passed since symptoms first appeared.
Minnesota	It used "people no longer need self-isolation" as a substitute for "recovered". However, it is reported that the definition has changed many times[2].
Montana	Persons who have cleared the illness and are released from isolation.
North Carolina	Estimates a median time to recovery of 14 days from the date of specimen collection for non-fatal COVID-19 cases who were not hospitalized, or if hospitalization status is unknown. The estimated median recovery time is 28 days from the date of specimen collection for hospitalized non-fatal COVID-19 cases.
North Dakota	Cases who tested positive and have since been released from isolation and are no longer contagious.
Oklahoma	Currently not hospitalized or deceased and 14 days after onset/report
Oregon	People are assumed to have recovered three days after resolution of all symptoms.
South Carolina	1) Those who reported being hospitalized/where hospitalization status was unknown: having no reported adverse outcome reported as of >32 days since their illness onset; 2) Those who reported not being hospitalized: having no reported adverse outcome reported as of >14 days since their illness onset.
Tennessee	1) have been confirmed to be asymptomatic by their local or regional health department and have completed their required isolation period or 2) are at least 21 days beyond the first test confirming their illness.
Texas	"Estimated based on several assumptions related to hospitalization rates and recovery times, which were informed by data available".

[2] https://www.startribune.com/coronavirus-covid-19-minnesota-tracker-map-county-data/568712601/

Utah	The number of cases whose first positive laboratory test was reported at least 21 days ago, excluding deaths.
Vermont	1) People who have tested positive for COVID-19 report they have recovered to our investigation teams during their follow-up calls. 2) Thirty days or more have passed since the date the person's illness began.
Wisconsin	the number of confirmed cases who are currently alive based on Wisconsin state vital records system data and had one or more of the following: 1) Documentation of resolved symptoms 2) Documentation of release from public health isolation 3) 30 days since symptom onset or diagnosis
Wyoming	1) A lab confirmed or probable case: resolution of fever without the use of fever-reducing medications and there is improvement in respiratory symptoms for 72 hours AND at least 10 days have passed since symptoms first appeared. 2) Cases with laboratory-confirmed COVID-19 who have not had any symptoms: at least 10 days have passed since the date of their first positive test and have had no subsequent illness provided they remain asymptomatic.

POPULATION COUNCIL
Ideas. Evidence. Impact.

Sources

[1] Commission on Social Determinants of Health, World Health Organization. Closing the gap in a generation: health equity through action on the social determinants of health. https://apps.who.int/iris/bitstream/handle/10665/43943/9789241563703_eng.pdf?sequence=1. Published 2008. Accessed June 10, 2020.

[2] Alder, N., Rehkopf, D. U.S. Disparities in Health: Descriptions, Causes, and Mechanisms. *Annu. Rev. of Public Health*. 2008;29:235-52. doi: 10.1146/annurev.publhealth.29.020907.090852.

[3] Centers for Disease Control. First Data Released on Maternal Mortality in Over a Decade. https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2020/202001_MMR.htm. Published January 30, 2020. Accessed June 10, 2020.

[4] Thebault, R., Tran, A., Williams, V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *The Washington Post*. April 7, 2020. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/?arc404=true. Accessed April 20, 2020.

[5] Kendi, I. Stop Blaming Black People for Dying of the Coronavirus. *The Atlantic.* April 14, 2020. https://www.theatlantic.com/ideas/archive/2020/04/race-and-blame/609946/. Accessed April 20, 2020.

[6]US: Address Impact of COVID-19 on Poor. Human Rights Watch. https://www.hrw.org/news/2020/03/19/us-address-impact-covid-19-poor#. Published March 19, 2020. Accessed April 20, 2020.

[7] Sex, gender, and COVID-19: overview and resources. Global Health 5050. https://globalhealth5050.org/covid19/. Accessed April 20, 2020.

[8] Centers for Disease Control. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019. https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm. Published April 17, 2020. Accessed April 20, 2020.

[9] Greenfieldboyce, N. How Is The Coronavirus Affecting Black Americans. NPR. https://www.npr.org/2020/04/08/829575332/how-is-the-coronavirus-affecting-black-americans. Published April 8, 2020. Accessed April 20, 2020.

[10] Quinn, s., Kumar, S. Health Inequities and Infectious Disease Epidemics: A Challenge for Global Health Security. *Biosecur Bioterror*. 2014 Sep 1; 12(5): 263–273. doi: 10.1089/bsp.2014.0032.

[11] Quinn, s., et al. Racial Disparities in Exposure, Susceptibility, and Access to Health Care in the US H1N1 Influenza Pandemic. *Am J Public Health*. 2011 February; 101(2): 285–293. doi: 10.2105/AJPH.2009.188029.

[12] HIV and African American Gay and Bisexual Men. Centers for Disease Control and Prevention. https://www.cdc.gov/hiv/group/msm/bmsm.html. Accessed April 20, 2020. Updated May 18, 2020.

[13] Klein SL, Jedlicka A, Pekosz A. The Xs and Y of immune responses to viral vaccines. Lancet InfectDis. 2010;10(5):338–49. doi: 10.1016/S1473-3099(10)70049-9.

[14] Racial Data Transparency. Johns Hopkins University & Medicine: Coronavirus Resource Center. https://coronavirus.jhu.edu/data/racial-data-transparency. Accessed April 20, 2020.

[15] Stafford, K., Hoyer, M., Morrison, A. Racial toll of virus grows even starker as more data emerges. *AP News.* https://apnews.com/8a3430dd37e7c44290c7621f5af96d6b. Published April 18, 2020. Accessed April 20, 2020.

[16] Centers for Disease Control. Coronavirus Disease 2019 (COVID-19). https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Accessed April 20, 2020.

[17] Johns Hopkins University & Medicine. Coronavirus Resource Center. COVID-19 United States Cases by County. https://coronavirus.jhu.edu/us-map. Accessed April 20, 2020.

[18] Centers for Disease Control. Coronavirus Disease 2019 Case Surveillance – United States, January 22-May 30, 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6924e2.htm?s_cid=mm6924e2_w. Accessed June 15, 2020.