# The Epidemiology of Diversity: COVID-19 Case Rate Patterns in California

Paul Hsu[1] · David E. Hayes-Bautista[2]

## Abstract

California's diverse population provides a natural laboratory for understanding how diseases and conditions interact within different racial/ethnic groups. This report seeks to illustrate the differential effects of the COVID-19 pandemic in the state's "majority-minority" population and to discuss the resulting implications for public health. Laboratory-confirmed COVID-19 cases in California (disaggregated by race/ethnicity into mutually exclusive groups) were integrated with their respective population values to create case rates per 100,000 population, categorized by age group and race/ethnicity. The case rates within each non-White population, in almost every age group, were higher than the White Non-Hispanic population, ranging from one-and-a-half to nearly six times as high. Public health prevention measures such as sheltering-at-home rely on standard assumptions and models. The disparity in case rates found here suggests that alternative narratives such as the epidemiology of diversity may inform additional policies or measures.

**Keywords** Epidemiology · Race/ethnicity · Covid-19 · Morbidity

## Introduction

California has been a "majority-minority" state for nearly two decades. In the 2018 American Community Survey (ACS), 62.5% of the state's population was what used to be called "minority": Latino, Black, Asian, Native Hawai'ian/Pacific Islander, or American Indian/Alaskan Native [1]. This diverse population provides a natural laboratory for understanding how diseases and conditions interact with different racial/ethnic (R/E) groups. Rather than start with the assumption that a region's epidemiological profile is driven by the non-Hispanic White population curve, with smaller groups varying from that norm, the epidemiology of diversity considers the profile of each group separately.

Almost from its beginning, the COVID-19 pandemic seems to have impacted some populations (*e.g.*, Black and Latino) more severely than other populations (*e.g.*, Non-Hispanic White). A recent JAMA "Viewpoint" written by a team from the National Institute on Minority Health and Health Disparities pointed out that "the pandemic presents a window of opportunity for achieving greater equity in the health care of all vulnerable populations." [2].

This brief report traces the pandemic's effects on different populations, using the COVID-19 case rates for six major R/E groups across different age groups.

## Methods

Data on California's cumulative number of laboratory-confirmed COVID-19 cases (136,191), as of June 9, 2020, were made available by the California Department of Public Health (CDPH) [3], disaggregated by race/ethnicity into exclusive groups: Latino (of any race), White Non-Hispanic (NH), Asian NH, African-American/Black NH, American Indian/Alaskan Native NH, Native Hawai'ian/Pacific Islander NH, Multi-racial NH, and Other NH. These categorizations are consistent with the guidance provided by the U.S. Census for establishing mutually

✉ Paul Hsu
  paulhsu@ucla.edu

1   Department of Epidemiology, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, 924 Westwood Blvd., Suite #200-R, Los Angeles, CA 90024, USA

2   Center for the Study of Latino Health and Culture, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, USA

exclusive groups. Hispanic, Spanish, or Latino ethnicity is identified first, regardless of race, followed by the other, non-Hispanic racial groupings [4, 5].

The CDPH compiles cases reported by local health departments, along with results from its own network of laboratories. Two types of diagnostics are available: molecular tests that detect the unique genetic material of SARS-CoV-2, and antigen tests that detect specific proteins found on the virus's surface. Antibody tests are also available, but these can only reveal the presence of a past infection, and are not recommended for diagnostic use [6].

Molecular tests using reverse transcription polymerase chain reaction (RT-PCR) are considered the gold standard [7]. California only uses PCR tests in its laboratories, so samples processed via the state have a relatively high degree of confidence. However, antigen tests may have been conducted by local health departments, leading to reported cases from such testing [8]. Data were also provided by age groups: 0−17, 18−34, 35−49, 50−64, 65−79, and 80+ .

This study derives population denominators for each race/ethnicity, disaggregated into these age groups, from the 2018 American Community Survey 1-year estimates, the most recent available [9]. Although the case counts are considered to be complete, there is potential variation in the population values. Employing the weights provided by ACS, 95% confidence intervals are calculated for each population and age stratum. The cases in each group are then divided by three values: the corresponding population figure, its upper bound, and its lower bound, resulting in rates with upper and lower limits per 100,000 population. Data are tabulated for all racial/ethnic categories except for Multi-Race and Other Non-Hispanic, but our figure

excludes American Indian/Alaska Native Non-Hispanic because of a relatively low number of cases.

All analyses are performed on publicly available, de-identified data, with no human subject participation; thus, the Office of the Human Research Protection Program (Institutional Review Board) considers this work exempt from review.

## Results

The epidemiology of diversity is shown in Fig. 1.

Overall, the curve for COVID-19 case rates observed in the White NH population in each age group from 0–17 to 80+ is consistently lower than the curve for all other R/E groups. The Asian NH curve starts out similar to the White NH in the age group 0–17, then quickly rises to about 50% higher than the White NH curve throughout the older age groups. The Black/African-American NH curve likewise starts out similar to White NH in the age group 0–17, then rises to about twice as high in the older adult groups. The Latino curve starts about three times as high as the White NH curve in the age group 0–17, and continues to be nearly three times as high throughout all six age groups, including the oldest adults. The Native Hawai'ian/Pacific Islander NH curve in the adult and older adult groups ranges from three to five times as high as the White NH curve. There are too few American Indian/Alaska Native NH cases in all age groups to calculate separate age-group rates, but where the cases approach n = 30, the rates trend higher than those for White NH.

Table 1 provides a summary of the case rate data, along with 95% confidence intervals, while Table 2 describes the distribution of missing age and race/ethnic information.

**Fig. 1** Case rates of COVID-19 by age group and race/ethnicity for California, 06-09-2020
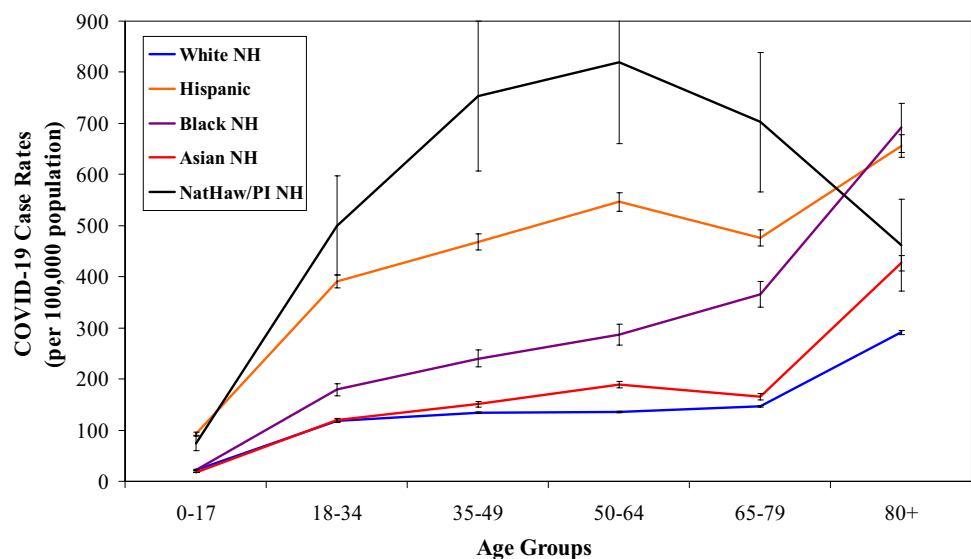
**Table 1** COVID-19 case rates per 100,000 population by age group and race/ethnicity: California, 2020

| | Children (ages 0–17) | | Young adults (ages 18–34) | | Early middle age (ages 35–49) | | Late middle age (ages 50–64) | | Older adults (ages 65–79) | | Oldest adults (ages 80+) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No | Rate (95% CI) | No | Rate (95% CI) | No | Rate (95% CI) | No | Rate (95% CI) | No | Rate (95% CI) | No | Rate (95% CI) |
| White NH | 487 | 21.31 (21.05, 21.57) | 3585 | 118.85 (117.68, 120.07) | 3589 | 134.02 (132.68, 135.39) | 4491 | 135.60 (134.57, 136.65) | 3472 | 146.29 (145.17, 147.46) | 2422 | 291.15 (287.54, 294.9) |
| Hispanic | 4346 | 92.79 (92.17, 93.41) | 16,951 | 391.04 (388.11, 394.02) | 14,840 | 468.25 (464.16, 472.39) | 12,046 | 546.71 (540.95, 552.68) | 4248 | 475.39 (467.35, 483.47) | 1678 | 655.90 (635.17, 677.61) |
| Black NH | 97 | 21.76 (21.09, 22.45) | 1037 | 178.94 (174.28, 183.65) | 1010 | 240.07 (232.67, 247.44) | 1249 | 286.96 (278.99, 295.29) | 832 | 366.32 (353.03, 379.87) | 484 | 691.45 (647.58, 739.1) |
| Asian NH | 172 | 17.44 (17.1, 17.8) | 1695 | 119.29 (117.35, 121.21) | 1977 | 151.06 (148.59, 153.57) | 2111 | 189.23 (186.08, 192.36) | 1120 | 165.70 (162.43, 169.14) | 934 | 426.56 (411.67, 441.99) |
| AI/AN NH | 8 | 41.44 (35.63, 49.48) | 56 | 176.38 (158.95, 197.32) | 53 | 210.32 (182.48, 241.14) | 41 | 135.79 (122.64, 152.47) | 20 | 114.55 (99.67, 130.12) | 13 | 313.03 (237.71, 421.4) |
| NH/PI NH | 19 | 73.43 (64.1, 84.62) | 172 | 499.81 (450.66, 566.96) | 213 | 753.69 (661.83, 847.96) | 198 | 819.60 (716.94, 939.44) | 86 | 702.73 (592.45, 839.3) | 20 | 461.47 (347.87, 620.12) |

Data for individuals identified as "2 or more races" (i.e., multi-race) and "Other" are not shown

*CI* confidence interval, *NH* non-Hispanic, *AI/AN* American Indian/Alaska Native, *NH/PI* Native Hawai'ian/Pacific Islander

**Table 2** Cases with missing race/ethnic information by age group: California, 2020

| | Total | Missing race/ethnicity | |
|---|---|---|---|
| | | Number | Percent (%) |
| Children (ages 0–17) | 8800 | 2547 | 28.9 |
| Young adults (ages 18–34) | 38,568 | 11,390 | 29.5 |
| Early middle age (ages 35–49) | 34,261 | 9634 | 28.1 |
| Late middle age (ages 50–64) | 30,914 | 8309 | 26.9 |
| Older adults (ages 65–79) | 14,907 | 4232 | 28.4 |
| Oldest adults (ages 80+) | 8562 | 2641 | 30.8 |
| No information on age | 179 | 102 | 57.0 |
| Total cases | 136,191 | 38,855 | 28.5 |

This table denotes the number of cases with missing race/ethnic information; it does not include 179 cases with missing age information

## Discussion

The major limitation of California's data was that, out of 136,191 cases reported, 28.5% (38,855) were missing data on race/ethnicity. The pattern of missing R/E information in each age group was relatively consistent, ranging from 26.9 to 30.8% (Table 2), and did not appear to be missing differentially. A complete count might change the specific rate values, but most likely would not significantly alter the patterns seen here. It is not likely that all of the cases with missing data would fall into one particular racial/ethnic group, nor is it likely that the missing cases would be perfectly apportioned among all groups.

One potential solution is to divide the cases with missing data across the different R/E groups based upon their R/E proportions in the population. Unfortunately, this alternative is limited in situations where groups are *dis*proportionately affected in relation to their numbers. We nevertheless examine the hypothetical results of such an approach: in almost all age groups, Latinos and other "minority" populations would have increased case rates, due to their larger numbers in younger cohorts, compared to White NH rates. Only in two age groups (65–79 and 80+) were White NH a majority (56% and 59%, respectively) [9]. White NH rates would increase in comparison to other groups, but not to the other groups' levels of severity (data not shown).

Case counts also depend on the availability and accessibility of testing, among other factors. Some cities and regions have not been able to acquire sufficient testing kits (availability); and even where these were available, not all individuals have had equal access to them, due to lack of transportation, awareness, symptoms, or other eligibility criteria for testing. Not only are these case numbers most likely an undercount of the actual scope of the disease, but the underreporting is almost certainly more pronounced in

groups with lower socioeconomic status and other barriers to testing, thus exacerbating these differential patterns.

While hospitalizations and death rates can also provide information, they are further "downstream" and may reflect additional structural factors, such as health care capacity. An "epidemic curve" of new cases over time would provide further context, reflecting the temporal course of the disease, but those data are not readily available by race/ethnicity. Similarly, "positivity" rates—the proportion of all individuals tested who had positive results—are not available by R/E group. Information on whether race/ethnicity was assigned or assumed by staff, or was a result of self-identification, is also not available. The limitations of using these categories have been well-documented [10], but they are necessary proxies for drawing attention to disparities between groups.

Our findings are consistent with reports in both popular media and published literature [11, 12]. In fact, the *New York Times* had to sue the Centers for Disease Control and Prevention (CDC) to gain access to the federal data. They concluded, "Black and Latino people have been disproportionately affected by the coronavirus in a widespread manner that spans the country, throughout hundreds of counties in urban, suburban and rural areas, and across all age groups." [11] (Note: only cases through the end of May, 2020, were included, and R/E identifiers were missing in more than half. The data included nearly 1.5 million infections at that time. The CDC acts as an aggregator for cases reported by state and territorial jurisdictions, so the California data used here is also the source of their California data.)

Other reliable databases include the Johns Hopkins Coronavirus Resource Center, which provides data on cases and deaths, and maps for the United States and worldwide, but no information on age or race/ethnicity, so a direct comparison is not possible [13]. Similarly, while the COVID Tracking Project at *The Atlantic* provides a comprehensive state-by-state dashboard, including a Racial Data Tracker, no age details are available [14]. California's records, analyzed in this report, provide the most granularity and are the data source for both the CDC and the Tracking Project. In fact, aggregate R/E counts (without age stratification) from the COVID Tracking Project for the relevant dates are exactly the same as those reported here.

Exposure to the coronavirus leads to the development of COVID-19 cases. A recent Economic Roundtable report notes that, in California, different industries and occupations may expose workers to the virus differently. The report notes that the state's labor force is strongly stratified by race/ethnicity, with some groups overrepresented in particular industries and occupations (*e.g.*, Latinos in farm work) that could expose them to the coronavirus more often than industries and occupations in which other groups predominate (*e.g.*, White NH in banking services) [15].

Additional factors may be responsible for the observed variation in rates. For example, Latinos continue to have the highest average household size, compared to other R/E groups [16]. Thus there is the potential for a family member who has been working outside the home to return and infect a greater number of individuals in the same household. The greater propensity of multigenerational households in minority (non-White) populations may also play a role, by potentially exposing those who are more vulnerable [17]. At the individual level, gender, smoking status, and co-morbid/underlying conditions have all been noted as increasing the risk of severe disease [18], but these factors do not necessarily lead to increased exposure.

California was one of the earliest states to introduce protective measures, in particular having non-essential workers shelter at home with their families. Those who had jobs that could be done from a distance tend to be salaried employees with health insurance coverage and regular sources of health care. While essential employees, such as physicians and nurses tending to COVID-19 cases in hospitals, are in theory provided with personal protective equipment (PPE), other essential workers are left out of the PPE mandate, although their work allows other families to shelter at home: farm workers growing food, workers shoulder-to-shoulder in meatpacking plants, grocery store checkout clerks, nursing home attendants, and non-professional hospital staff, such as cleaning and maintenance personnel.

These occupations and industries are largely filled by Latinos, Blacks, Asians, and other minority populations, and we suggest that this job-related exposure most likely explains the differential patterns described in this report.

## Implications for Public Health

Public health interventions are based on our models of public health threats. The standard epidemiological assumption that the White NH rate is the norm, with minority groups varying from that norm, has led to widespread application of the shelter-at-home intervention. The epidemiology of diversity could have provided a different starting point for coronavirus exposure interventions, by pointing out that the vast majority of "essential" workers are not medical professionals, but individuals who perform manual labor and provide essential goods and services to those sheltering at home—more often than not without the benefit of PPE, much less health insurance or access to a regular source of health care.

We therefore urge a greater awareness and understanding of the numerous, and sometimes implicit, factors that inform our public health models. For example, one of the recommended non-pharmaceutical interventions during the spread of a novel infectious disease is to shelter at home.

But for large swaths of the population, sheltering at home is not feasible, either from an economic perspective or simply from the requirements of their jobs. Consideration of alternative narratives might have led to additional policies specific to these vulnerable groups, for example, paid leave or sick leave beyond the amount mandated by the CARES Act, or additional PPE requirements.

Recent events have only reinforced the reality that social and physical inequities persist and are pervasive in society. Public health has a responsibility to address these disparities, in order to form a more equitable community, one with physical, mental, and social well-being for all [19].

## References

1. California Department of Finance. 2014−2018 American Community Survey (5-year estimates), race and Hispanic/Latino, California, counties, incorporated cities, and census designated places. http://www.dof.ca.gov/Reports/Demographic_Reports/American_Community_Survey. Accessed 25 May 2020.
2. Webb Hooper M, Nápoles AM, Perez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA. 2020;323(24):2466–7. https://doi.org/10.1001/jama.2020.8598.
3. California Department of Public Health. COVID-19 Race and Ethnicity Data (2020). https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/Race-Ethnicity.aspx. Accessed 10 June 2020.
4. United States Census Bureau. About Race. https://www.census.gov/topics/population/race/about.html. Accessed 27 Sept 2020.
5. United States Census Bureau. Race. https://www.census.gov/quickfacts/fact/note/US/RHI625219. Accessed 27 Sept 2020.
6. U.S. Food and Drug Administration. Coronavirus Testing Basics. https://www.fda.gov/consumers/consumer-updates/coronavirus-testing-basics. Accessed 26 Sept 2020.
7. Guglielmi G. The explosion of new coronavirus tests that could help to end the pandemic. Nature. 2020;583:506–9. https://doi.org/10.1038/d41586-020-02140-8.
8. California Coronavirus COVID-19 Testing Task Force. https://testing.covid19.ca.gov/. Accessed 26 Sept 2020.
9. Ruggles S, Flood S, Goeken R, et al. IPUMS USA: Version 10.0 [dataset]. Minneapolis: IPUMS; 2020. https://doi.org/10.18128/D010.V10.0.
10. Yudell M, Roberts D, DeSalle R, Tishkoff S. Taking race out of human genetics. Science. 2016;351(6273):564–5. https://doi.org/10.1126/science.aac4951.
11. Oppel Jr RA, Gebeloff R, Lai KKR, Wright W, Smith M. The fullest look yet at the racial inequity of coronavirus. The New York Times, July 5, 2020. https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html. Accessed 27 Sept 2020.
12. Van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet. 2020;395(10232):1243–4. https://doi.org/10.1016/S0140-6736(20)30893-X.
13. Johns Hopkins Coronavirus Resource Center. US Map FAQ. https://coronavirus.jhu.edu/us-map-faq. Accessed 27 Sept 2020.

14. The COVID Tracking Project at The Atlantic. About the racial data tracker. https://covidtracking.com/race/about. Accessed 27 Sept 2020.

15. Flaming D, Burns P. In Harm's way: California workers at high risk of unemployment in the COVID-19 pandemic. economic roundtable. April 2020. https://economicrt.org/publication/in-harms-way Accessed 25 May 2020.

16. United States Census Bureau. America's families and living arrangements: 2016, Table AVG1. https://www.census.gov/data/tables/2016/demo/families/cps-2016.html. Accessed 27 Sept 2020.

17. Pew Research Center. A record 64 million Americans live in multigenerational households, 2018. https://www.pewresearch.org/fact-tank/2018/04/05/a-record-64-million-americans-live-in-multi generational-households/. Accessed 27 Sept 2020.

18. Centers for Disease Control and Prevention. People with certain medical conditions. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html. Accessed 27 Sept 2020.

19. Preamble to the Constitution of WHO as adopted by the international health conference, New York, 19 June–22 July 1946; signed 22 July 1946 (Official Records of WHO, no. 2, p. 100); entered into force 7 April 1948.