

Disproportionate Life Expectancy Decline Between Non-White and White Americans During the COVID-19 Pandemic*

Russell Luchin and Michael Shmelev

February 20, 2024

This study replicates the findings of Nuzzo and Ledesma (2023) through the lens of racial stratification. Findings indicate a strong link between differences in racialized healthcare causing a comparatively larger decline in life expectancy in racialized communities during the COVID-19 pandemic. Study is conducted by cross-referencing replicated graphs with each other and alongside prior studies on the same topic. Suggestions for improving racial stratification in healthcare are discussed.

Table of contents

1 Introduction	2
2 Data	3
2.1 Methodology	3
2.2 Raw Data	3
2.3 Prepared Data	4
2.4 Measurement	5
3 Results	6
3.1 Life Expectancy	6
3.2 Change in Life Expectancy (2019-2020 and 2020-2021)	7
3.3 Expected vs Observed Death Rate	8

*Code, data, and .QMD available at: <https://github.com/rlichin/Disproportionate-Dcline-In-Life-Expectancy-USA-2020>. All code and analysis done with R programming language (R Core Team, 2023). Credit to Jennifer B. Nuzzo and Jorge R. Ledesma for the replication package and the original paper that was used as a basis for this project.

3.4 Cross Referencing the Life Expectancy Change and Death Rate	8
4 Discussion	9
4.1 Racial Disparities in Life Expectancy	9
4.2 Expected vs Observed Death Rate	10
4.3 Weaknesses in Analysis	10
4.4 R and R Library Acknowledgment	11
References	11

1 Introduction

Nuzzo and Ledesma (2023) explore the link between American pandemic preparedness and the United States' disproportionately high COVID-19 mortality rate. To this end, Nuzzo and Ledesma (2023) discuss in depth the shortcomings of the United States' COVID-19 response in comparison to similarly prepared allies. In this paper, we replicate key figures from Nuzzo and Ledesma (2023) and adapt them to explore how America's racial stratification bleeds into healthcare access across communities. This paper aims to evaluate the difference in impact of the COVID-19 pandemic on average life expectancy between non-white American communities and white American communities. The estimand of this study is the differential effect of the pandemic on the life expectancy across racial groups.

This paper finds that the COVID-19 pandemic had a disproportionate effect on non-white groups when compared to white groups of Americans. We find that racialized communities, when averaged, had more than double the decrease in life expectancy when compared to white communities in the United States. Furthermore, this paper explores the potential links between life expectancy decrease during the COVID-19 pandemic and the difference in observed vs expected death rate per 1000 people. Analysis finds that average life expectancy decrease is similar to observed death rate increase over expected, but does not find a conclusive link.

The findings of this paper supplement the findings of Nuzzo and Ledesma (2023) alongside others (Dwyer-Lindgren et al. 2022), complementing existing studies which explore the racial stratification of the healthcare system. There is no shortage of studies which highlight the extreme racial stratification of the American healthcare system and how it affects racialized communities negatively, and this paper adds to the growing repertoire of studies on this topic. The findings in this paper aim to inform relevant parties of the necessity to mend racial stratification in the American healthcare system by any means, as the disproportionate affect of the COVID-19 pandemic on non-white groups continue to demonstrate that it is a matter of life and death.

2 Data

2.1 Methodology

Data was acquired from Nuzzo and Ledesma, 2023. Raw data was imported from Arias et al, 2022 and The World Bank, 2022. This data was acquired from a replication package for Nuzzo and Ledesma, 2023. The data was cleaned and replicated as per the instructions of Nuzzo and Ledesma, 2023 and the associated replication package.

Data used to create Figures 1 and 2 were further cleaned and aggregated to isolate American life expectancy by ethnic group, separated between two groups “Non-Hispanic White” and “Non-White”. As the raw data separates ethnic groups per the US Census (Hispanic, Non-Hispanic AIAN, Non-Hispanic Black, Non-Hispanic Asian, Non-Hispanic White) all groups that were outside the “Non-Hispanic White” category were averaged by summarizing their life expectancy at birth into a single variable. “Non-Hispanic White” was untouched beyond the initial preparation from Nuzzo and Ledesma, 2023. This data originates from the Center for Disease Control (Arias et al. 2022).

Figure 3 was isolated from Figure 4(a) from Nuzzo and Ledesma, 2023 to highlight expected American mortality rate against observed American mortality rate (all causes). This data comes from The World Bank (Bank 2022) and is cleaned according to the replication package from Nuzzo and Ledesma, 2023. Other than isolating the American mortality rate from a larger figure, the data is left untouched and is accurate to the cleaned data retrieved from Nuzzo and Ledesma, 2023.

2.2 Raw Data

Table 1: Life Expectancy by Racial Group

year_id	All	Hispanic	Non-Hispanic AIAN	Non-Hispanic Asian	Non-Hispanic Black	Non-Hispanic White
2006	77.8	80.3	NA	NA	73.1	78.2
2007	77.9	80.9	NA	NA	73.2	78.2
2008	78.1	81.0	NA	NA	73.7	78.4
2009	78.5	81.2	NA	NA	74.2	78.7
2010	78.7	81.4	NA	NA	74.7	78.8
2011	78.7	81.6	NA	NA	74.9	78.8
2012	78.8	81.9	NA	NA	75.1	78.9
2013	78.8	81.9	NA	NA	75.1	78.8
2014	78.9	82.1	NA	NA	75.3	78.8
2015	78.7	81.9	NA	NA	75.1	78.7
2016	78.7	81.8	NA	NA	74.9	78.6

year_id	All	Hispanic	Non-Hispanic AIAN	Non-Hispanic Asian	Non-Hispanic Black	Non-Hispanic White
2017	78.6	81.8	NA	NA	74.9	78.5
2018	78.7	81.8	NA	NA	74.7	78.6
2019	78.8	81.9	71.8	85.6	74.8	78.8
2020	77.0	77.9	67.1	83.6	71.5	77.4
2021	76.1	77.7	65.2	83.5	70.8	76.4

Table 1 prints the raw data from the data-frame “life”, created from the raw data in “life_table.csv” acquired from Arias et al, 2022. This data shows average life expectancy, per year from 2006 to 2021, separated by racial/ethnic group.

Table 2: All Cause Death Rates

Country Name	2020	2021	2022
Aruba	9.690000	11.235000	NA
Africa Eastern and Southern	7.620548	8.036899	NA
Afghanistan	7.113000	7.344000	NA
Africa Western and Central	10.601297	10.678952	NA
Angola	7.821000	8.009000	NA
Albania	10.785000	11.325000	NA

Table 2 shows raw data from the data-table “dths”, created from raw data in “all_cause_dth_rates.csv” acquired from The World Bank, 2022. This table shows the first column of the table (country name) alongside the last 3 columns of the total table (all cause death rate per 1000 people, separated by year) for the first 6 rows of the total data-table.

2.3 Prepared Data

Table 3: Life Expectancy by Racial Group (Cleaned)

year_id	Racial-ethnic group	life_exp
2006	Non-Hispanic White	78.200
2006	Non-White	76.700
2007	Non-Hispanic White	78.200
2020	Non-White	75.025
2021	Non-Hispanic White	76.400
2021	Non-White	74.300

Table 3 shows the first and last 3 rows of the prepared life expectancy data-frame. All racial groups outside of “Non-Hispanic White” have been aggregated into a single variable per year from 2006, while “Non-Hispanic White” remains untouched from the raw data.

Table 4: All Cause Death Rate (Prepared)

location_name	iso3	year	dth_rate	year_id	t	preds
United States	USA	year_2013	8.215	2013	1	8.214250
United States	USA	year_2014	8.237	2014	2	8.304786
United States	USA	year_2015	8.440	2015	3	8.395321
United States	USA	year_2016	8.493	2016	4	8.485857
United States	USA	year_2017	8.638	2017	5	8.576393
United States	USA	year_2018	8.678	2018	6	8.666929
United States	USA	year_2019	8.700	2019	7	8.757464
United States	USA	year_2020	10.300	2020	8	8.848000
United States	USA	year_2021	10.400	2021	9	8.938536

Table 4 shows the prepared data from “dths”, from “all_cause_death_rates.csv”. In the preparation of this data, the same regression used to predict all cause death rates from Nuzzo and Ledesma, 2023 was used to create the “preds” column. This regression used to compute expected death rates comes from running country specific linear regressions, as described and Nuzzo and Ledesma, 2023.

2.4 Measurement

Data was measured through 4 independent variables. These variables are expected death rate, observed death rate, life expectancy, and total change in life expectancy. The latter 2 variables were separated by race, between “Non-White” and “Non-Hispanic White” in our prepared data. Measured data was nominal, collected and categorized by racial/ethnic group, year, and # of deaths. Linear regression was used to calculate expected death rate, using the linear regression replicated from Nuzzo and Ledesma, 2023.

Quantitative data was collected to create these variables, with Nuzzo and Ledesma, 2023 being the foundational source for the measured variables. Measured data was prepared as per the replication package from Nuzzo and Ledesma, 2023, with variations to the “life” and “dths” data-frames to track racial stratification more closely. This was done as described in Section 2.3.

3 Results

3.1 Life Expectancy

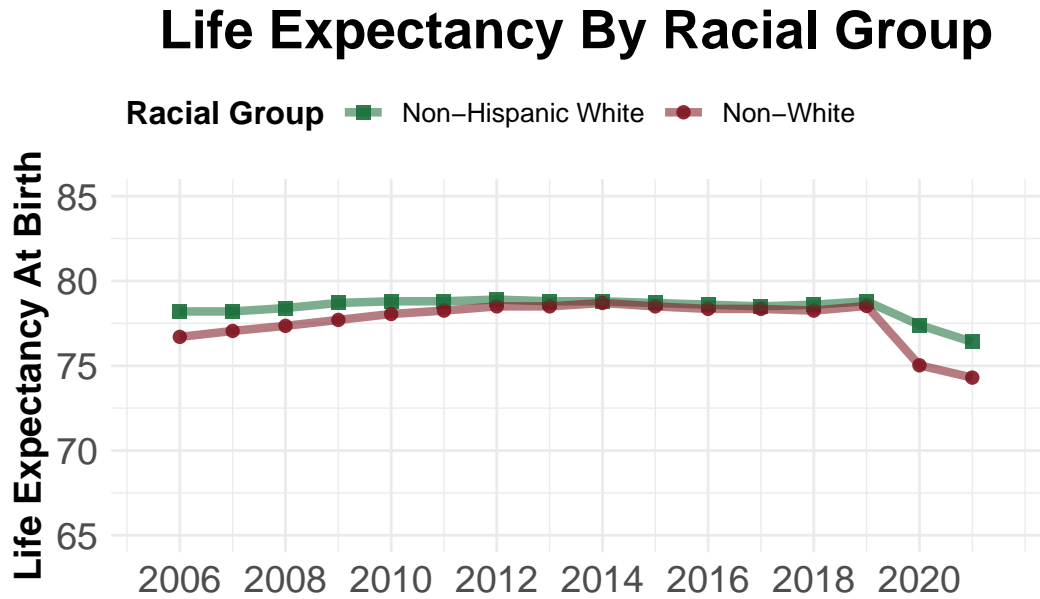


Figure 1 plots our prepared life-expectancy data-frame we are able to identify significant trends in American life expectancy regarding racial disparity. Between 2006 and 2012 we observe a gradual closing of life expectancy disparities between white and non-white Americans. From 2012 to 2019 we observe life expectancy to be not significantly different between white and non-white Americans. This trend could be due to certain low-income healthcare programs being implemented in 2010, such as Obamacare (Public Affairs (ASPA) 2022). However, following the COVID-19 pandemic in 2020, we observe a dramatic decrease in life expectancy across both our measured categories. White Americans observe about half the decrease from non-white Americans, who see a far more dramatic decline in average life expectancy than white Americans.

It should also be noted that both racial groups see a reversal in trends from 2006 to 2019 following the COVID-19 pandemic. From 2020 onwards, we observe a slight decline in average life expectancy when there seemed to be a steady increase up until 2019. This could point to a greater failure of the American healthcare system, although data for this is inconclusive as there is very little to extrapolate from 2020 onwards in this dataset.

3.2 Change in Life Expectancy (2019-2020 and 2020-2021)

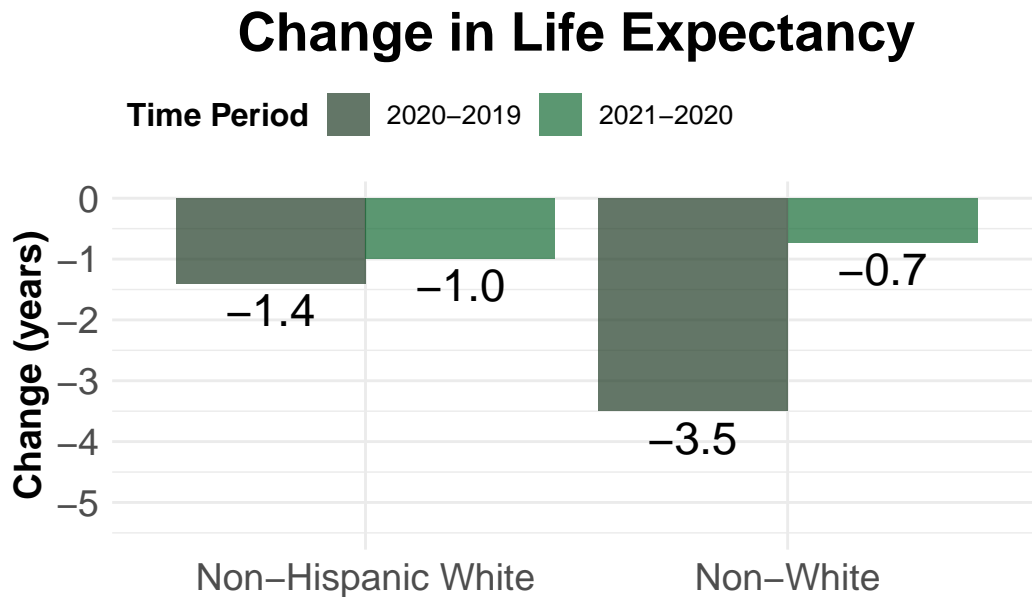


Figure 2 graphs change in life expectancy between non-white and white Americans, observing a disparity. While average life expectancy only dropped by about a year for white Americans, non-white Americans observed nearly a 4 year cut (averaged) to their life expectancy. This is a disparity of over 200% when, prior to the COVID-19, observed life expectancy was minimally different between white and non-white Americans.

A disparity this large is abnormal, considering the gradual trend of life expectancy becoming equal between white and non-white Americans up to 2019 (Figure 1). Figure 2 supplements the dramatic difference in life expectancy dropping between white and non-white Americans by giving a nominal value (difference in life expectancy) to these groups. A difference of around 200% implies great stratification between white and non-white Americans, in spite of regular efforts to normalize healthcare access Smith (2021).

3.3 Expected vs Observed Death Rate

USA Expected Death Rate

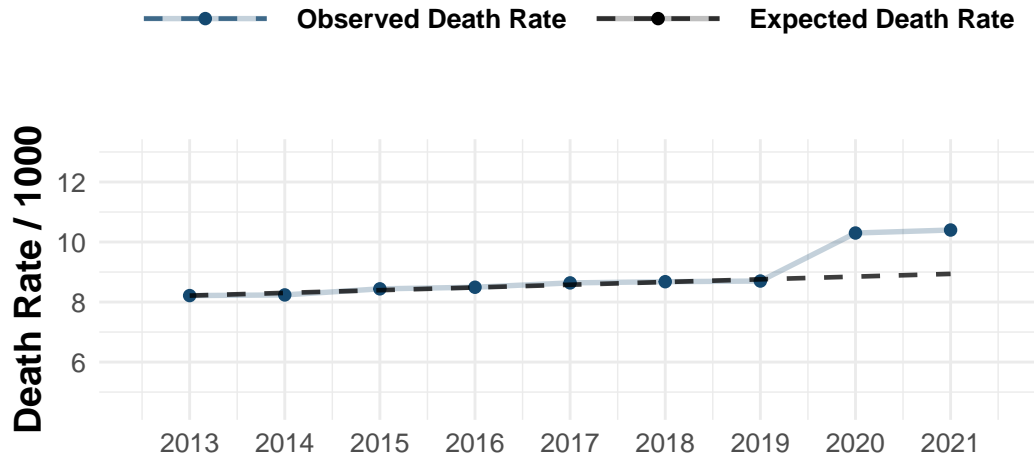


Figure 3 graphs our expected death rate alongside our observed death rate. As seen in Figure 3, there is a substantial difference between our predicted death rate (calculated using linear regression with death rates from 2013-2019) and our observed death rate; particularly in 2020.

What should be noted is that, in 2020, the COVID-19 pandemic struck the globe as nations struggled to mitigate the damage the infection caused. The linear regression from Nuzzo and Ledesma, 2023 has no way to account for this, and therefore the expected death rate continues the trends observed between 2013-2019.

Cross referencing this data with Figure 1, we note some consistent data points. When comparing average life expectancy from 2013 to 2019, we note a slight decline in the expected lifespan across both white and non-white Americans. Figure 3 visualizes a slight increase in death rate, which is consistent with our findings in Figure 1.

3.4 Cross Referencing the Life Expectancy Change and Death Rate

Comparing Figures 2 and 3, we observe a potentially significant correlation. Figure 3 observes a significant difference between expected death rate and observed death rate, the difference being 2~ deaths per 1000 people. By averaging the change in life expectancy from Figure 2, we can reach a similar value. We average the change in life expectancy with this equation: **A**

$= (x+y)/2$, where A is the average of X and Y , X is the average life expectancy decrease of non-white Americans, and Y the decrease of white Americans.

Focusing on the change between 2019 and 2020, we observe an average life expectancy drop of 2.5. Cross-referencing this average with the difference between expected and observed death rate from Figure 3, we note a similar increase of about 2.2~ more deaths than expected. We see a similar pattern when observing the change between 2020 and 2021, where the average decrease in life expectancy is .85 years and the increase in death rate is about .5 years.

As these values are not identical, it is a possibility that this is a coincidence. However, it can be further hypothesized that, if there was a lower disparity between the drop in average life expectancy between white and non-white Americans, then the observed increases in death rate per 1000 would be lower. This is because the drop observed in Figure 2, and the climb observed in Figure 3, have very similar nominal values (differences of .2-.3 when an average is calculate across groups in Figure 2 and compared to Figure 3).

4 Discussion

4.1 Racial Disparities in Life Expectancy

Racial disparities in life expectancy are not a novel discovery, and Figures 1 and 2 complement findings from wider-reaching studies such as *A Systematic Analysis of Health Disparities*(Dwyer-Lindgren et al. 2022). The reasons for these disparities are widespread, ranging from lack of healthcare access in indigenous communities (Dwyer-Lindgren et al. 2022) to long histories of healthcare discrimination towards African-Americans (Reyes 2020).

There have been attempts by the American federal government to mend this disparities, such as the implementation of “Obamacare” (Public Affairs (ASPA) 2022) in 2010. While the implementation of Obamacare correlates with a gradual normalization between non-white and white communities in terms of life expectancy (Figure 1), the underlying social factors which lead to racial discrimination in healthcare remain widespread (Dwyer-Lindgren et al. 2022). Obamacare allowed low-income communities to access healthcare in greater proportions, but how could it help low-income communities that don’t have access to modern healthcare facilities altogether? Racial disparities in life expectancy encompass more than discrimination; it encompasses decades of system racial discrimination in conjunction with a lack of access to healthcare (Dwyer-Lindgren et al. 2022) and/or distrust of medical institutions (Sekimitsu et al. 2022).

While the large discrepancy between white communities and non-white ones in their respective life expectancy drops during COVID are concerning, they are far from surprising. Much work needs to be done to mend these barriers, and not all of them can be fixed with legislation allowing more Americans to afford healthcare. There needs to be a widespread, social overhaul

of how the medical system treats non-white Americans. Anything from easing the historical trauma found in affected groups (Sekimitsu et al. 2022), to increasing healthcare access in indigenous communities (Smith 2021), will work to mend this severe stratification we observe in Figures 1 and 2.

4.2 Expected vs Observed Death Rate

In Nuzzo and Ledesma, 2023 the primary focus is to elaborate on why the United States had a substantially higher COVID death rate as opposed to similarly prepared nations. Figure 3 is an isolated graph of America's expected vs observed death rate during the peak of the COVID-19 pandemic. Nuzzo and Ledesma, 2023 discusses how a likely cause in this disparity is America's weakness in pandemic preparedness to their allies. The focus of Nuzzo and Ledesma, 2023 centers around America's all around weakness in preparedness and response, citing procurement, healthcare capacity, and legislative action regarding the COVID-19 pandemic. Despite the strengths of this paper, it neglects to cover in-depth another factor to why the United States had a comparatively high observed death rate to its expected: healthcare access disparity.

Figures 1 and 2 highlight the severe differences in non-white life expectancies to white life expectancies. When these graphs are cross referenced, we observe a very similar average life expectancy decrease to the difference in expected vs observed death rate. While this can be considered a coincidence, it would be hard to deny that a differences in healthcare access across racial groups contributed to the high discrepancy in the death rate from Figure 3. For example, African American communities had a higher distrust of vaccines than other racial groups (Sekimitsu et al. 2022). When compared to the data Nuzzo and Ledesma, 2023, we observe further how African Americans had one of the highest drops in life expectancy among the measured racial groups.

The disparity in the death rate in Figure 3 only tells part of a story. It's only when we link everything together that we start to see a story of historic and systemic racial disparities in access to healthcare. The COVID-19 pandemic merely inflamed underlying differences in healthcare Sekimitsu et al. (2022), rather than being the sole contributor to this disparate life expectancies.

4.3 Weaknesses in Analysis

While Figures 1-3 provided interesting insight to racial stratification in American healthcare, the data is weakened by being adapted from Nuzzo and Ledesma, 2023. This paper exists as a replication before an analysis, and therefore could benefit from a wider range of data which was collected for the purposes of studying racial stratification in healthcare. Rather, this paper adapts data from Nuzzo and Ledesma, 2023 to study this topic. This is not to say that the data is inaccurate, or that Nuzzo and Ledesma, 2023 is inadequate; rather it serves

to point out that this study would benefit from a wider, more specialized, dataset to analyze racial stratification.

Futhermore, Figures 2 and 3 could be linked by using a model to see how observed death rate and average life expectancy decrease interact. While a conclusive link is outside the scope of this paper, it isn't unrealistic to see a potential connection between the figures in question. Creating a dataframe to measure life expectancy drop alongside death rate increase is a potential approach which future researchers can explore to study this potential link further.

4.4 R and R Library Acknowledgment

The analysis conducted in this paper would not be possible without the R programming language (R Core Team 2023) or the accompanying libraries. Thank you to the creators of ggpubr (Kassambara 2023), ggplot2 (Wickham 2016), readxl (Wickham and Bryan 2023), lubridate (Grolemund and Wickham 2011), data.table (Barrett et al. 2024), dplyr (Wickham et al. 2023), and knitr (Xie 2023) for making this paper possible.

References

- Arias, Elizabeth, Kenneth D. Kochanek, Jiaquan Xu, and Betzaida Tejada-Vera. 2022. "Provisional Life Expectancy Estimates for 2022." *Vital Statistics Rapid Release*.
- Bank, The World. 2022. "Death Rate, Crude (Per 1,000 People) - 2022."
- Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, and Toby Hocking. 2024. *Data.table: Extension of 'Data.frame'*. <https://CRAN.R-project.org/package=data.table>.
- Dwyer-Lindgren, Laura, Parkes Kendrick, Yekaterina O Kelly, Dillon O Sylte, Chris Schmidt, Brigitte F Blacker, Farah Daoud, et al. 2022. "Life Expectancy by County, Race, and Ethnicity in the USA, 2000–19: A Systematic Analysis of Health Disparities." *The Lancet* 400 (10345).
- Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with Lubridate." *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Kassambara, Alboukadel. 2023. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr>.
- Public Affairs (ASPA), Assistant Secretary for. 2022. "About the Affordable Care Act."
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reyes, Maritza Vasquez. 2020. "The Disproportional Impact of COVID-19 on African Americans." *Health and Human Rights* 22 (2).
- Sekimitsu, Sayuri, Jessica Simon, Mika Matsuuchi Lindsley, Melissa Jones, Umin Jalloh, Titi-layo Mabogunje, Jordyn Kerr, et al. 2022. "The Disproportional Impact of COVID-19 on

- African Americans.” *American Journal of Health Promotion* 36 (8). <https://doi.org/10.1177/08901171221099270>.
- Smith, Mary. 2021. “Native Americans: A Crisis in Health Equity.” *Human Rights Magazine*, no. 3.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, and Jennifer Bryan. 2023. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.