

Population density and basic reproductive number of COVID-19 across United States counties

Karla Therese L. Sy,¹ Laura F. White,¹ Brooke Nichols^{1,2*}

¹ Boston University School of Public Health, Boston, Massachusetts, USA

² University of the Witwatersrand, Johannesburg, South Africa

***Correspondence should be addressed to:** Brooke Nichols, brooken@bu.edu

801 Massachusetts Avenue Crosstown Center, Department of Global Health, 3rd Floor
Boston, MA 02118

Article Summary Line: U.S. counties with greater population density have greater rates of transmission of SARS-CoV-2, likely due to increased contact rates in areas with greater density.

Running Title: Population density and R_0 of COVID-19 across US counties

Keywords: Population density; Basic reproductive number; COVID-19; SARS-CoV-2; Disease transmission; United States; Spatial

Abstract

The basic reproductive number (R_0) is a function of contact rates among individuals, transmission probability, and duration of infectiousness. We sought to determine the association between population density and R_0 of SARS-CoV-2 across U.S. counties, and whether population density could be used as a proxy for contact rates. We conducted a cross-sectional analysis using linear mixed models with random intercept and fixed slopes to assess the association of population density and R_0 . We also assessed whether this association was differential across county-level main mode of transportation-to-work percentage. Counties with greater population density have greater rates of transmission of SARS-CoV-2, likely due to increased contact rates in areas with greater density. The effect of population density and R_0 was not modified by private transportation use. Differential R_0 by population density can assist in more accurate predictions of the rate of spread of SARS-CoV-2 in areas that do not yet have active cases.

Introduction

The COVID-19 pandemic has infected millions of people globally, and there are over 400 thousand reported deaths and 7 million confirmed cases of COVID-19 worldwide.¹ Transmission of airborne and directly transmitted pathogens,²⁻⁴ such as SARS-CoV-2 (the causative agent of COVID-19), have been previously shown to be density-dependent. Population density facilitates transmission of disease via close person-to-person contact,⁵⁻⁸ and may support sustained disease transmission due to increased contact rates.⁹⁻¹¹ Large urban areas have more opportunities for disease transmission, and hotspots of SARS-CoV-2 have been mostly concentrated in cities.¹²

The basic reproductive number (R_0) describes the contagiousness and transmissibility of pathogens, and is a function of contact rates among individuals, transmission probability, and number of infective individuals.¹³ Thus, R_0 estimates of COVID-19 are not exclusively determined by the pathogen, and variability in R_0 depends on local sociobehavioral and environmental settings, including population density.¹² During the initial phase of the outbreak, or the exponential growth period, we hypothesize that spatial heterogeneity in R_0 occurs in part due to geographic variability in contact rates, since transmission probability and population size remain constant. Since the exponential growth period occurs prior to the implementation of non-pharmaceutical interventions (NPIs), such as face coverings and social distancing, we would expect that the probability of transmission per contact would be the same equal across settings. Moreover, contact networks are also affected by transportation systems that facilitate disease spread due to increased interconnectivity and mobility between different geographic areas;^{14,15} thus, we also hypothesize that the association of population density and R_0 may be differential

depending on transportation accessibility, and areas that lack access to efficient modes of transportation would not have the same SARS-CoV-2 growth rate, even in high density areas.

In the current COVID-19 pandemic, data-driven analyses of the association of population density and disease transmission has not been systematically quantified. The estimation of differential R_0 using these area-level factors can assist in more accurate predictions of the rate of spread of SARS-CoV-2 in geographic settings where cases have only begun to rise. In this study, we examine the association of population density with R_0 of COVID-19 across United States counties.

Methods

Data

We obtained publicly available daily COVID-19 case and death data among United States counties from the New York Times.¹⁶ For each county, we assumed that the exponential growth period was one week prior to the second daily increase in cases. We assumed that the period of exponential growth approximately lasted 18 days, as previous research have shown the COVID-19 exponential period to be around 20-24 days in New York City,¹⁷ and calibrated it accordingly to create reasonable curves that approximated exponential growth across the counties (**Supplemental Materials Appendix 1**). The algorithm ensured that the virus had taken hold in the area and allowed a sufficient number of days to estimate the exponential growth rate. We restricted calculation of R_0 to counties with greater than 25 cases at the end of the exponential growth period, as R_0

cannot be estimated accurately with sparse data and it would be uncertain if the county was experiencing a sustained outbreak with community transmission. Data on the primary mode of transportation to work and median household income were obtained from the most recent 5-year American Community Survey (ACS) 2014-2018 survey estimates from the United States Census Bureau.¹⁸ Population and land area were obtained from the 2010 census, and density was calculated by population divided by total square km. All census data were extracted using the R package *tidycensus*.¹⁹

Statistical analyses

We first compared the densities of counties included in the final analytical sample to those that did not have sufficient case counts with a two-sample Wilcoxon test. R_0 was calculated using the Wallinga and Lipsitch method.²⁰ We then conducted a cross-sectional analysis using linear mixed models with random intercept for each county and fixed slopes to assess the association of population density and R_0 . The models controlled for state-level effects using random intercepts, and county-level main mode of transportation to work percentage. We also adjusted for median household income to control for any potential confounding between the association of private car ownership and R_0 .

We fit 4 models with R_0 as the outcome and the following factors as covariates: Model 1: population density; Model 2: population density and the percent of individuals reporting private transportation as their main mode of transportation to work; Model 3: population density and the percent of individuals reporting private transportation, and

median household income; Model 4: population density, percent of individuals reporting private transportation, median household income, and the interaction of private transportation use with population density.

Sensitivity analyses

We conducted three sensitivity analyses to address the limitations of our approach and assess the robustness of our results. First, we conducted the main analysis using death counts to estimate R_0 to limit bias due to differential availability of testing by geographic location. We used the same exponential period as the cases, but with a lag of 14 days to account for the delays from symptom onset to deaths of cases.^{21,22} Moreover, the analysis of deaths was restricted to counties with greater than 10 deaths and more than 5 daily increases in incident deaths, in order to appropriately estimate R_0 in counties with sufficient death counts. Second, we excluded counties within a radius of 15 miles, the average commuting miles in the United States,²³ from counties with densities greater than the 75th percentile. Removing these adjacent counties would demonstrate the extent of biases due to individuals commuting from surrounding counties to cities. If cases are imported from more densely populated (i.e. cities) to less dense counties, we could potentially be biasing our estimates downwards. Lastly, we conducted an analysis excluding influential counties with a Cook's distance measure over $4/N$ for each model, in order to ensure that findings were not driven by influential data points.

All analyses were conducted in R version 4.0.0.²⁴ The figure and removal of adjacent counties in the sensitivity analyses were done with ArcGIS.²⁵

Results

The United States has 3,221 counties and county equivalents. When restricting to counties with greater than 25 cases, 1,151 (35.73%) counties were included (**Figure 1**). The median density in counties included and not included in the analysis were 59.8 people/km² (IQR: 23.64-150.47) and 11.2 people/km² (IQR: 3.65-23.64) respectively, and the difference was statistically significant ($p < 0.0001$). The median R_0 among the counties was 1.66 (IQR: 1.34-2.11).

An increase in one unit of log population density increased R_0 by 0.16 (95% CI=0.13 to 0.19) (Model 1; **Table 1**), or the doubling of population density increased the R_0 on average by 0.11 (95% CI=0.09 to 0.13). When adjusted for percent of private transportation and median household income, the association of log population density and R_0 remained unchanged (Model 3; **Table 1**). There was no significant interaction, and the effect of population density on R_0 was the same among counties with a larger percentage with private cars as their transportation to work (Model 4; **Table 1**). R_0 decreased by 0.12 (95% CI= -0.02 to -0.04) with an 10% increase in private transportation as the main commute mode, accounting for population density and median household income (Model 3; **Table 1**).

In all three sensitivity analyses, population density remained positively associated with R_0 , demonstrating the robustness of our main analysis. First, death data was used to calculate R_0 from 310 counties. The median R_0 among the counties that had sufficient death counts was 1.40 (IQR: 1.05-1.78). The unadjusted association between population density and R_0 remained consistent ($\beta=0.18$, 95% CI=0.14 to 0.23) (Model 1a; **Table 2**), and there were no significant interactions (Model 4a; **Table 2**). Next, there were 288

counties above the 75th percentile, and 414 counties that were within 15 miles of these counties high-density counties. We removed these 414 counties from the sample, and using the subsample of 737 counties, our findings remained consistent (**Table 2**). Influential counties were also not driving the association of population density and R_0 , and our results remained robust (**Table 2**). For the two sensitivity analyses excluding counties adjacent to highly dense counties and excluding high influence counties, however, the association of private transportation usage and R_0 did not remain (Models 2a, 3a, 2b, 3b; **Table 2**).

Discussion

Our findings show that the basic reproductive number (R_0) is associated with population density, even when percent of individuals that use private transportation and median income were accounted for. In these settings, greater population density may potentially facilitate interactions between susceptible and infectious individuals in densely-population networks, which sustain continued transmission and spread of COVID-19. Moreover, we see that population density continues to have an important impact on disease transmission regardless of transportation accessibility and median income, suggesting that the opportunity for effective contacts are mostly driven by crowding in denser areas, increasing the contact rates necessary for disease spread. However, we did not see that density-dependence is differential across transportation accessibility. Even though transmission is less in lower density areas (i.e. rural areas), rural settings may eventually disproportionately be more vulnerable to COVID-19 morbidity and mortality. Individuals in rural areas are generally older, have more

underlying conditions, have less access to care, and have fewer ICU beds, ventilators, and facilities needed for severe COVID-19 treatment.²⁶⁻²⁸ Further research is needed on the overall burden of COVID-19 across the spectrum of population density.

Geographic estimates of R_0 of SARS-CoV-2 need to take into account the specific area's population density, since the R_0 estimate is dependent on both the pathogenicity of the virus as well as environmental influences. In countries where cases are only starting to climb, such as countries in Latin America and Southern Africa,^{1,29} or there is a resurgence of cases, such as India, Iraq, and Israel,³⁰ area-specific density can assist in predictions of R_0 , which is important because epidemiological forecasts and predictive models are sensitive to small changes in R_0 inputs. Accurate estimation of R_0 consequently lead to more precise estimates of the epidemic size, so that governments can appropriately allocate resources and coordinate mitigation strategies. Moreover, as cities and states reopen in the United States, and if there is a second-wave of infections, areas with higher density accessibility will likely have greater SARS-CoV-2 resurgence.

Our study has a number of limitations. While we demonstrate that population density is associated with R_0 , we estimated R_0 based on the number of reported cases; therefore, the incidence of COVID-19 across US counties may be underestimated at varying rates due to differential testing. Testing data at the county-level currently do not exist, and we were unable to adjust for the number of tests performed. To mitigate this limitation, we included a random intercept term to adjust for state-level effects, and thus differential testing across states were accounted by our model. Differential testing by local governments within states are less likely to strongly impact our findings, as most funding and budgets for COVID-19 is distributed at the state-level.^{31,32} We also conducted a

sensitivity analysis using death data which demonstrates the robustness of our findings. Additionally, we had to limit our analysis to counties that had sufficient case data in order to accurately estimate R_0 . Given our findings that the counties excluded in the analysis had a significantly lower density and presumably very low R_0 due to lack of cases, the true association between population density and R_0 would likely be greater than what we report in our analysis. Another limitation is that our model also assumes homogenous mixing, which may can be an oversimplification of the heterogeneity in contact patterns within populations.^{4,33} However, previous research has shown that population structure only changes R_0 estimates slightly,³⁴ and assumptions of well-mixed populations are valid in small-to-medium spatial scales.³⁵ Moreover, our method loses spatial granularity in assessing R_0 in counties, especially in counties with spatially heterogenous clustering. The aim of our study, however, was to provide a generalizable estimate of the association between population density and R_0 , in order to appropriately estimate potential for disease transmission, rather than a microspatial estimate that may not be generalizable to other settings. Finally, an important confounder that we were unable to adjust for is the number of importations of SARS-CoV-2 in these counties, as more urbanized areas are more likely to have links with countries and other states where the virus could have originated from. Even so, we still see that once an area is seeded with COVID-19, the growth rate is greater in denser areas during the time period prior to implementation of NPIs.

In summary, counties with greater population density have greater rates of transmission of SARS-CoV-2, likely due to increased contact rates in areas with greater population density. Population density affects the network of contacts necessary for

disease transmission, and SARS-CoV-2 R_0 estimates need to consider this variability for proper planning and resource allocation, particularly as epidemics newly emerge and old epidemics resurge.

Author contributions: KTLS, BEN contributed to conceptualization. KTLS contributed to data acquisition. KTLS, LFW, and BEN contributed to data analysis. All authors contributed to interpretation of results and manuscript writing.

Sources of funding: KTLS and BEN were funded for this work by United States Agency for International Development (USAID) through the following cooperative agreement: AID-OAA-A-15-00070. LFW was supported by NIH R01 GM122876. The funding bodies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. All authors have seen and approved the manuscript.

Disclaimers: The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Competing interests: The authors have declared no conflicts of interest.

Tables and Figures

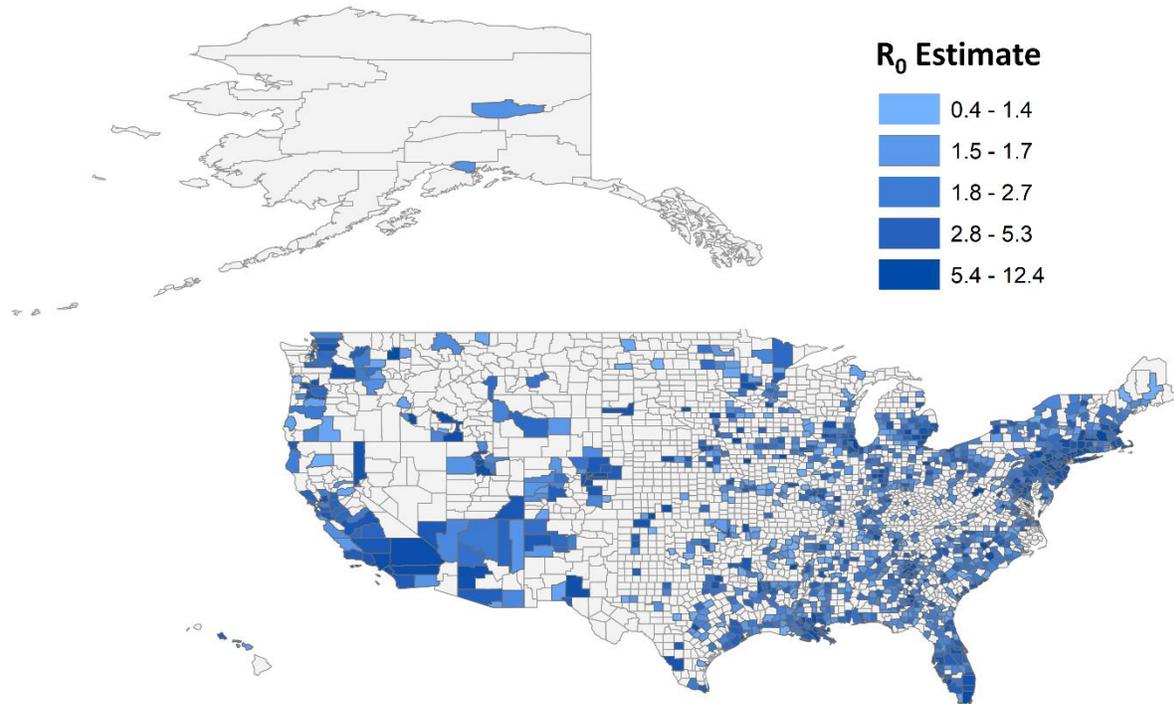


Figure 1. Basic reproductive number (R_0) estimates across United States counties.

Larger R_0 indicates greater transmission during the initial phase of the outbreak, or the exponential growth period. We restricted calculation of R_0 to counties with greater than 25 cases at the end of the exponential growth period ($n=1,151$), as R_0 cannot be estimated accurately with sparse data and it would be uncertain if the county was experiencing a sustained outbreak with community transmission.

Table 1. Linear mixed models (random intercept, fixed slope) evaluating the association between population density and basic reproductive number (R_0) among United States counties. Estimates for each model is a slope (beta) with a null of 0; a positive slope indicates that an increase in the log of population density increases R_0 by the beta estimate for the log of population density. The interaction term indicates that the association of population density and R_0 differs depending on the percentage of people using the private transportation for work.

	Model 1	Model 2	Model 3	Model 4
Log of population density	0.16*** (0.13-0.19)	0.15*** (0.12, 0.18)	0.15*** (0.12, 0.18)	0.2* (-0.06, 0.47)
Percent of private transportation (10%)		-0.12*** (-0.19, -0.04)	-0.12** (-0.2, -0.04)	-0.08 (-0.27, 0.11)
Median household income (\$10,000)			0 (-0.03, 0.03)	0 (-0.03, 0.03)
Interaction of population density and transportation				-0.01 (-0.04, 0.02)

pvalue * < 0.05; ** < 0.01; *** < 0.001

Model 1- unadjusted association of population density and R_0

Model 2 – association of population density adjusted for the percent of individuals reporting private transportation

Model 3 – association of population density, percent of individuals reporting private transportation, and median household income

Model 4 – association of population density, percent of individuals reporting private transportation, median household income, and the interaction of population density and transportation

Table 2. Sensitivity analysis of linear mixed models (random intercept, fixed slope) using (a) deaths only, (b) removing counties within 15 miles of high density counties, and (c) removing high influence counties

	Model 1	Model 2	Model 3	Model 4
Deaths only				
Log of population density	0.18*** (0.14, 0.23)	0.14*** (0.09, 0.19)	0.12*** (0.07, 0.17)	0.48* (0.01, 0.94)
Percent of private transportation (10%)		-0.16** (-0.26, -0.06)	-0.15** (-0.24, -0.05)	0.14 (-0.24, 0.53)
Median household income (\$10,000)			0.06** (0.03, 0.1)	0.07*** (0.03, 0.11)
Interaction of population density and transportation				-0.04 (-0.09, 0.01)
No counties adjacent to high density counties				
Log of population density	0.17*** (0.14, 0.2)	0.17*** (0.13, 0.2)	0.16*** (0.13, 0.2)	0.02 (-0.13, 0.17)
Percent of private transportation (10%)		0.01 (-0.07, 0.1)	-0.01 (-0.12, 0.1)	-0.12 (-0.28, 0.03)
Median household income (\$10,000)			0.02 (-0.03, 0.06)	0.01 (-0.03, 0.05)
Interaction of population density and transportation				0.03 (0, 0.06)
Removing high influence counties				
Log of population density	0.18*** (0.15, 0.2)	0.18*** (0.16, 0.21)	0.18*** (0.15, 0.2)	0.38* (0.04, 0.72)
Percent of private transportation (10%)		-0.01 (-0.09, 0.07)	-0.02 (-0.1, 0.07)	0.09 (-0.1, 0.29)
Median household income (\$10,000)			0 (-0.03, 0.02)	0 (-0.02, 0.03)
Interaction of population density and transportation				-0.02 (-0.06, 0.02)

pvalue * < 0.05; ** < 0.01; *** < 0.001

Model 1- unadjusted association of population density and R0

Model 2 – association of population density adjusted for the percent of individuals reporting private transportation

Model 3 – association of population density, percent of individuals reporting private transportation, and median household income

References

1. Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard. <https://coronavirus.jhu.edu/map.html>.
2. van Boven M, Koopmans M, Du Ry van Beest Holle M, et al. Detecting Emerging Transmissibility of Avian Influenza Virus in Human Households. *PLOS Computational Biology*. 2007;3(7):e145.
3. McCallum H, Barlow N, Hone J. How should pathogen transmission be modelled? *Trends in Ecology & Evolution*. 2001;16(6):295-300.
4. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*. 2013;244(2):125-134.
5. Wu T, Perrings C, Kinzig A, Collins JP, Minter BA, Daszak P. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. *Ambio*. 2017;46(1):18-29.
6. Wu X, Tian H, Zhou S, Chen L, Xu B. Impact of global change on transmission of human infectious diseases. *Science China Earth sciences*. 2014;57(2):189-203.
7. Lienhardt C. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiologic reviews*. 2001;23(2):288-301.
8. Rader B, Scarpino S, Nande A, et al. Crowding and the epidemic intensity of COVID-19 transmission. *medRxiv*. 2020:2020.2004.2015.20064980.
9. Miller JC. Spread of infectious disease through clustered populations. *Journal of the Royal Society, Interface*. 2009;6(41):1121-1134.
10. Mei S, Chen B, Zhu Y, Lees MH, Boukhanovsky AV, Sloot PMA. Simulating city-level airborne infectious diseases. *Computers, Environment and Urban Systems*. 2015;51:97-105.
11. Smieszek T, Fiebig L, Scholz RW. Models of epidemics: when contact repetition and clustering should be included. *Theoretical biology & medical modelling*. 2009;6:11.
12. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number ($R(0)$). *Emerg Infect Dis*. 2019;25(1):1-4.

13. Anderson RM, May RM. *Infectious Diseases of Humans, Dynamics and Control*. OUP Oxford; 1992.
14. Pinter-Wollman N, Jelić A, Wells NM. The impact of the built environment on health behaviours and disease transmission in social systems. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1753):20170245.
15. Bian L. Spatial Approaches to Modeling Dispersion of Communicable Diseases – A Review. *Transactions in GIS*. 2013;17(1):1-17.
16. New York Times. Coronavirus (Covid-19) Data in the United States. 2020. <https://github.com/nytimes/covid-19-data/blob/master/us-counties.csv>.
17. Sy KTL, Martinez ME, Rader B, White LF. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. *medRxiv*. 2020:2020.2005.2028.20115949.
18. United States Census Bureau. 2012-2016 American Community Survey (ACS) 5-year Estimates. 2020; <https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2016/5-year.html>.
19. Walker K. tidy census: Load US Census Boundary and Attribute. Data as 'tidyverse' and 'sf'-Ready Data Frames. R package version 0.9.5. 2020.
20. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings Biological sciences*. 2007;274(1609):599-604.
21. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious diseases*. 2020:S1473-3099(1420)30243-30247.
22. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *The Lancet Infectious diseases*. 2020.
23. United States Department of Health. Bureau of Transportation Statistics. Omnibus Household Survey. 2003.
24. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing [computer program]. Vienna, Austria 2017
25. *ArcGIS Desktop: Release 10* [computer program]. Redlands, CA: Environmental Systems Research Institute; 2011.

26. Casey M, Evenson A, Moscovice I, Wu Z. Availability of Respiratory Care Services in Critical Access and Rural Hospitals. 2018.
<https://rhrc.umn.edu/publication/respiratory-care-services-in-critical-access-and-rural-hospitals>. Accessed July 6, 2020.
27. Ranscombe P. Rural areas at risk during COVID-19 pandemic. *The Lancet Infectious diseases*. 2020;20(5):545.
28. Souch JM, Cossman JS. A Commentary on Rural-Urban Disparities in COVID-19 Testing Rates per 100,000 and Risk Factors. *J Rural Health*. 2020:10.1111/jrh.12450.
29. Rivers M, Gallón N, Pedroso R, Rahim Z. Mexico and parts of Brazil reopen after lockdown -- despite surging coronavirus cases 2020.
30. India Loosens Restrictions, Despite Coronavirus Surge. In: Times TNY, ed2020.
31. U.S. Department of Health & Human Services. HHS Delivers Funding to Expand Testing Capacity for States, Territories, Tribes. 2020.
<https://www.hhs.gov/about/news/2020/05/18/hhs-delivers-funding-to-expand-testing-capacity-for-states-territories-tribes.html#:~:text=HHS%20Delivers%20Funding%20to%20Expand%20Testing%20Capacity%20for%20States%2C%20Territories,support%20testing%20for%20COVID%2D19>.
32. U.S. Department of Treasury. Coronavirus Relief Fund: Guidance for State, Territorial, Local, and Tribal Governments 2020.
<https://home.treasury.gov/system/files/136/Coronavirus-Relief-Fund-Guidance-for-State-Territorial-Local-and-Tribal-Governments.pdf>.
33. Borremans B, Reijnders J, Hens N, Leirs H. The shape of the contact and density function matters when modelling parasite transmission in fluctuating populations. *Royal Society Open Science*. 2017;4(11):171308.
34. Trapman P, Ball F, Dhersin J-S, Tran VC, Wallinga J, Britton T. Inferring R0 in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface*. 2016;13(121):20160288.
35. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *Journal of travel medicine*. 2020;27(3).