

# Time-varying associations between COVID-19 case incidence and community-level sociodemographic, occupational, environmental, and mobility risk factors in Massachusetts

**Koen Tieskens**

Boston University School of Public Health

**Prasad Patil**

Boston University School of Public Health

**Jonathan I. Levy**

Boston University School of Public Health

**Paige Brochu**

Boston University School of Public Health

**Kevin J. Lane**

Boston University School of Public Health

**M. Patricia Fabian**

Boston University School of Public Health

**Fei Carnes**

Boston University School of Public Health

**Beth M. Haley**

Boston University School of Public Health

**Keith R. Spangler**

Boston University School of Public Health

**Jessica H. Leibler** (✉ [jleibler@bu.edu](mailto:jleibler@bu.edu))

Boston University School of Public Health

---

## Research Article

**Keywords:** COVID-19, risk factors, mitigation efforts, Coronavirus

**Posted Date:** February 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-237622/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Infectious Diseases on July 16th, 2021. See the published version at <https://doi.org/10.1186/s12879-021-06389-w>.

## Abstract

**Background:** Associations between community-level risk factors and COVID-19 incidence are used to identify vulnerable subpopulations and target interventions, but the variability of these associations over time remains largely unknown. We evaluated variability in the associations between community-level predictors and COVID-19 case incidence in 351 cities and towns in Massachusetts from March to October 2020.

**Methods:** Using publicly available sociodemographic, occupational, environmental, and mobility datasets, we developed mixed-effect, adjusted Poisson regression models to depict associations between these variables and town-level COVID-19 case incidence data across five distinct time periods. We examined town-level demographic variables, including z-scores of percent Black, Latinx, over 80 years and undergraduate students, as well as factors related to occupation, housing density, economic vulnerability, air pollution (PM<sub>2.5</sub>), and institutional facilities.

**Results:** Associations between key predictor variables and town-level incidence varied across the five time periods. We observed reductions over time in the association with percentage Black residents (IRR=1.12 CI=(1.12-1.13) in spring, IRR=1.01 CI=(1.00-1.01) in fall). The association with number of long-term care facility beds per capita also decreased over time (IRR=1.28 CI=(1.26-1.31) in spring, IRR=1.07 CI=(1.05-1.09) in fall). Controlling for other factors, towns with higher percentages of essential workers experienced elevated incidence of COVID-19 throughout the pandemic (e.g., IRR=1.30 CI=(1.27-1.33) in spring, IRR=1.20, CI=(1.17-1.22) in fall). Towns with higher percentages of Latinx residents also had sustained elevated incidence over time (e.g., IRR=1.19 CI=(1.18-1.21) in spring, IRR=1.14 CI=(1.13-1.15) in fall).

**Conclusions:** Town-level COVID-19 risk factors vary with time. In Massachusetts, racial (but not ethnic) disparities in COVID-19 incidence have decreased over time, perhaps indicating greater success in risk mitigation in selected communities. Our approach can be used to evaluate effectiveness of public health interventions and target specific mitigation efforts on the community level.

## Background

As of January 2021, the United States had the highest absolute number of coronavirus disease 2019 (COVID-19) cases and deaths in the world.<sup>1</sup> Within the US, disease incidence has varied substantially among states with different policy interventions and adherence to public health guidance,<sup>2-4</sup> and there is also significant variability within states.<sup>5</sup> For example in Massachusetts, during the two-week period from January 10–23, 2021, COVID-19 average daily incidence exceeded 100 confirmed cases per 100,000 persons in multiple urban communities (including Chelsea, Lawrence and New Bedford), with a low of zero in a number of more rural communities.<sup>6</sup>

Early in the pandemic, multiple community-level factors were associated with higher COVID-19 incidence rates, with disproportionate burdens among communities with more racial and ethnic diversity and workers in essential services.<sup>7</sup> As the pandemic evolved over time, factors associated with elevated case incidence are likely to have evolved as well, given changes in work patterns, behaviors, and state and local policies. To our knowledge, no studies to date have investigated the time-dependent association of COVID-19 incidence rate at a town level with key predictors.

In this study, we present a novel set of models analyzing associations between publicly available town-level data and COVID-19 incidence, with the goal of evaluating changes in key town-level predictors across Massachusetts over time. We ran parallel regression models with the same set of predictors over five different time periods from the first recorded case in Massachusetts in March 2020 to October 2020 to assess the temporal variability of the relative magnitude and significance of these predictors on COVID-19 incidence. Characterization of the time-variant nature of these predictors may assist in future efforts to target interventions to specific subpopulations and assess the effectiveness of mitigation strategies on the community level over time.

## Methods

Time periods: We used publicly available case incidence data for 351 cities and towns published by the Massachusetts Department of Public Health (MA DPH) from April 14, 2020 through October 29, 2020.<sup>8</sup> Case data were centrally reported to MA DPH by all laboratories across the state that conducted COVID-19 testing, and cumulative case counts and rates published weekly. We stratified that case data across 5 time periods selected to reflect distinct waves of case incidence in the state, starting with the spring “first wave” (March 2 – April 14, 2020, and April 15 – June 3), the summer nadir (June 4 – July 15, 2020 and July 16 – September 2, 2020), and the early months of the “second wave” in the fall (September 3 – October 29, 2020). The specific end date was chosen based on COVID-19 case incidence data availability at the time of study.

Data sources: Table 1 lists the sociodemographic, occupational, environmental, and mobility datasets at the city/town level which were integrated with data on COVID-19 incidence during the five time periods. Sociodemographic, occupational and economic data were extracted from the 2014–2018 American Community Survey (ACS) 5-year estimates at the census tract level. Data were scaled to town level by downloading ACS data in absolute numbers, aggregating to town level, and proportions calculated with ACS town population estimates.<sup>9</sup> Some rural census tracts contained several smaller towns, thus these had identical ACS percentage values. Estimates for percentage of essential workers were modified from ACS data based on an integrated dataset developed by the American Civil Liberties Union, which restricted ACS data on service workers to those job types considered “essential” during the pandemic including healthcare practitioners, transportation occupation, food preparation, etc.<sup>10</sup> Data on town-level imprisoned population were obtained through MassGIS<sup>11</sup> and long-term care beds by town were obtained from MA DPH.<sup>8,12,13</sup> The percentage of people commuting to work was estimated using the SafeGraph Social Distancing Metrics dataset derived from anonymized cell phone mobility data, where a commuter is defined as someone who spends 8 hours a day M-F in a single location away from their home block group.<sup>14</sup> Annual PM<sub>2.5</sub> concentrations were obtained from Kloog et al. at 1 km grid resolution, averaged by town for the year of 2015, and linked to reflect a general measure of air pollution.<sup>15</sup> Population density, as a proxy for local transmission risk, was calculated per town using 100 meter cell size population grids derived from the 2010 Census counts available at the census block level (assuming equal spatial distribution of people within census blocks). For each town we averaged the cell values depicting population totals in each cell at different Euclidean distance radii (5, 20, 50, and 100 km) to capture the effects of urbanicity and population density at different scales.<sup>16</sup>

Table 1

Sociodemographic, occupational, environmental, and mobility risk factors used to predict COVID-19 case incidence in cities and towns of Massachusetts

Variable name	Description
% Racial minorities	Percentage of non-white people <sup>a</sup>
% Black	Percentage of people identifying as African American/ Black <sup>a</sup>
% Hispanic/Latinx	Percentage of people identifying as Latino/Hispanic <sup>a</sup>
% Age > 80	Percentage of people older than 80 years <sup>a</sup>
% Age > 70	Percentage of people older than 70 years <sup>a</sup>
% Age < 20	Percentage of people younger than 20 years <sup>a</sup>
% Below poverty	Percentage of people living below poverty line <sup>a</sup>
% Disabilities	Percentage of people with disability <sup>a</sup>
% Essential services	Percentage of people with essential service job <sup>a</sup>
% No health insurance	Percentage of people without health insurance coverage <sup>a</sup>
% More than 2 per room	Percentage of housing units with occupancy > 2 persons per room <sup>a</sup>
% More than 1.5 per room	Percentage of housing units with occupancy > 1.5 persons per room <sup>a</sup>
% More than 1 per room	Percentage of housing units with occupancy > 1 persons per room <sup>a</sup>
% Public transportation	Percentage of people traveling to work by public transportation <sup>a</sup>
% Undergraduates	Percentage of people enrolled in undergraduate higher education <sup>a</sup>
Prison population	Percentage of incarcerated population <sup>b</sup>
LTCF (# beds)	Number of long term care facility beds as percentage of population <sup>c</sup>
PM <sub>2.5</sub>	Mean annual PM <sub>2.5</sub> level <sup>d</sup>
Population density (5km)	Number of people living within 5km radius from town <sup>a</sup>
Population density (20km)	Number of people living within 20km radius from town <sup>a</sup>
Population density (50km)	Number of people living within 50km radius from town <sup>a</sup>
Population density (100km)	Number of people living within 100km radius from town <sup>a</sup>

<sup>a</sup> Data derived from the 2014–2018 American Community Survey<sup>9</sup>

<sup>b</sup> Data derived from MassGIS<sup>11</sup>

<sup>c</sup> Data derived from the Massachusetts Department of Public Health<sup>8</sup>

<sup>d</sup> Data derived from Kloog et al. 2014<sup>15</sup>

<sup>e</sup> Data derived from SafeGraph mobility datasets<sup>14</sup>

Variable name	Description
% Going to work	Estimated percentage of population going to work outside home <sup>e</sup>
<sup>a</sup>	Data derived from the 2014–2018 American Community Survey <sup>9</sup>
<sup>b</sup>	Data derived from MassGIS <sup>11</sup>
<sup>c</sup>	Data derived from the Massachusetts Department of Public Health <sup>8</sup>
<sup>d</sup>	Data derived from Kloog et al. 2014 <sup>15</sup>
<sup>e</sup>	Data derived from SafeGraph mobility datasets <sup>14</sup>

Statistical Analysis: We developed a series of mixed-effect Poisson regression models to predict COVID-19 case incidence by town for each of the five distinct time periods of the pandemic. We used the most recent population estimate<sup>17</sup> of each town as an offset to account for differences in population per town. To minimize multi-collinearity while maintaining a good fit, we used a backward selection process of covariates where we stepwise excluded covariates with a variance inflation factor higher than 2.5 in the regression of the first time period.<sup>18</sup> All numerical predictors were normalized to zero mean and standard deviation of 1. We included the county of each town as a random effect to control for spatial autocorrelation of residuals (351 towns divided over 14 counties). For every period except the first period ending on April 14, we included an ordinal variable of the COVID-19 case count per capita of the previous period, divided into quintiles, as a random effect to adjust for between-period temporal autocorrelation. To assess collinearity structure of the predictor variables, we calculated Pearson correlations between all bivariate predictor combinations and presented them in a correlation matrix. Mixed-effect models were executed in R 3.6.1<sup>19</sup> using the *glmer* function from the *lme4* package.<sup>20</sup>

## Results

The backward predictor selection process resulted in a set of eight fixed-effect predictor variables that had a good fit in all 5 phases of the pandemic between the beginning of March and the end of October 2020. Excluded variables either caused high levels of multi-collinearity or had no significant association with the response variable. The correlation matrix in Fig. 1 shows a cluster of population density variables with public transportation and PM<sub>2.5</sub> in the upper left corner and a cluster of housing crowdedness with percent of the population working in essential services, under the poverty line, or with disabilities. Both clusters show positive correlations with % Black population and Hispanic/Latino (henceforth referred to as Latinx) population. Point estimates and associated 95% confidence intervals are presented for all eight predictors in the model across all time periods (Table 2).

Table 2

Significant associations between sociodemographic and economic predictors and COVID-19 case incidence by time period

Variable	Incidence rate ratio (95% CI) <sup>a</sup>				
	Time 1 <sup>b</sup>	Time 2	Time 3	Time 4	Time 5
% Age > 80 <sup>c</sup>	1.04 (1.02–1.07)	1.14 (1.12–1.16)	1.03 (0.98–1.08)	1.05 (1.01–1.09)	1.03 (1-1.05)
% Black	1.12 (1.12–1.13)	1.1 (1.09–1.1)	1.06 (1.04–1.07)	1.05 (1.04–1.06)	1.01 (1-1.01)
% Hispanic/Latinx	1.19 (1.18–1.21)	1.13 (1.12–1.14)	1.15 (1.13–1.17)	1.07 (1.06–1.09)	1.14 (1.13–1.15)
% No health insurance	0.85 (0.84–0.87)	1.09 (1.08–1.11)	1.14 (1.09–1.18)	1.21 (1.18–1.25)	1.06 (1.04–1.08)
% Essential services	1.3 (1.27–1.33)	1.21 (1.19–1.23)	1.2 (1.15–1.25)	1.24 (1.2–1.28)	1.2 (1.17–1.22)
LTCF beds	1.28 (1.26–1.31)	1.23 (1.21–1.24)	1.09 (1.05–1.13)	1.1 (1.07–1.13)	1.07 (1.05–1.09)
Population (20km)	1.21 (1.19–1.23)	0.98 (0.97–0.99)	1.01 (0.98–1.05)	1.23 (1.2–1.26)	0.98 (0.96-1)
% Undergraduates	1.02 (1.01–1.04)	0.98 (0.97–0.99)	0.96 (0.93-1)	0.97 (0.95–0.99)	1.06 (1.04–1.07)
<sup>a</sup> Incidence rate ratios generated through mixed-effect multivariable Poisson regression. Models were adjusted for all other covariates presented here.					
<sup>b</sup> Dates for study periods: Time 1: before April 14; Time 2: April 15 – June 3; Time 3: June 4 – July 15; Time 4: July 16 – September 2; Time 5: September 3 – October 29					
<sup>c</sup> Variables reflect percent of the town population by given characteristic, so percent of town population > 80 years or the percent of the town population Black/African American, for example. Data generated from the 2014–2018 American Community Survey. <sup>9</sup>					

Of the final selection of predictors, almost all showed substantial change in association with COVID-19 cases over the study period (Fig. 2). Whereas an increase of one standard deviation in the percentage of town residents identifying as Black was associated with an increase of more than 10% in COVID-19 cases before April 14, the association steadily decreased and became statistically non-significant by September-October. At the start of the pandemic, we observed a nearly 20% increase in COVID-19 cases by town associated with one standard deviation increase in the percentage of town residents identifying as Latinx, but the strength of this positive association varied over time, with a decreasing association from April to September and an uptick in the association in the last phase of our model (September 3rd to October 29th).

We observed a decreasing trend in the association between percent of the town population older than 80 years and the number of long term care facility beds per town with COVID-19 incidence, indicating that the presence of older residents and institutional care facilities in a town had a stronger association with incidence in the early, but not the later, periods of the study. The percent of essential workers by town consistently showed a significant positive association with COVID-19 incidence throughout all time periods.

Population density, measured as the total number of people living within 20 km from the boundary of each town, showed notable changes in association with COVID-19 case incidence, with high positive associations in the first and fourth time periods, and non-significant associations during the other time periods. The undergraduate student population, measured as

the percentage of the town population enrolled in undergraduate education, was associated with a small positive effect on COVID-19 incidence early in the pandemic and a larger association after September 3rd, when several universities throughout the state resumed with in-person or hybrid classes after shutting down for in-person learning in March/April 2020.

## Discussion

Our findings illustrate the time-varying nature of sociodemographic and economic predictors of COVID-19 case incidence at the community level in Massachusetts during an eight-month period. These observations suggest that fixed assumptions regarding community-level COVID-19 vulnerability, such as increased risk among Black communities or continued elevated relative risk among the elderly, may not accurately represent the pandemic at any given point in time, and that these associations should be continually reassessed though time for relevance alongside shifting mitigation efforts, policies and individual behavior. While other studies have identified racial/ethnic disparities in COVID-19 incidence in Massachusetts and the extent to which these patterns are explained by social factors,<sup>7</sup> our study provides novel insight about how both patterns and predictors change over time. Likewise, associations that remain significant over time with consistent coefficients in these adjusted models, such as workers in essential services, may suggest that existing approaches to reduce risk within these subgroups are less effective, perhaps due to structural challenges in reducing certain exposures.

We note that the community-level predictor variables used here serve as proxies for underlying factors that influence individual SARS-CoV-2 exposure risk. This distinction is important in our analysis, as it remains possible that analyses using individual-level predictors would lead to different findings. However, our analysis provides insight into sociodemographic patterns of COVID-19 and subpopulations who may be at elevated risk, informing community-scale public health interventions, and vaccination strategies by location, as well as a useful and adaptable structure to assess risk alongside public data.

The consistently elevated risk observed in communities with increased Latinx populations (in models adjusted for essential workers and other sociodemographic variables) may underscore challenges in reaching these communities with successful interventions, or barriers to reduced exposure in these communities. The sustained elevated risk of Latinx populations throughout the first eight months of the pandemic is consistent with recent studies.<sup>7,21</sup> In our models, Latinx population was positively correlated with greater housing density, suggesting that this ethnicity variable may serve in part as a proxy for crowded housing (> 1 person/room) in our models. Within-household transmission is an established risk factor for COVID-19 transmission,<sup>22,23</sup> and it is possible that our findings here reflect established challenges in reducing transmission within housing environments where residents are unable to meaningfully distance from an infectious individual.

We did not observe a correlation between housing density and Black population in a town, or between this race variable and other sociodemographic covariates in our model. This finding suggests that other factors beyond the scope of our analysis may be responsible for the elevated COVID-19 risks faced by towns with higher percentage of Black residents, especially early in the pandemic. Our findings parallel core conclusions of Figueroa et al., who evaluated cross-sectional COVID-19 incidence alongside demographic data in Massachusetts from March-May 2020.<sup>7</sup> The authors noted racial disparity in disease incidence in the early wave after adjustment for essential workers, immigration, and household size, while the association between COVID-19 cases and Latinx population was attenuated in models adjusted for these factors. Together, our studies support the hypotheses that systemic racism and inequities not otherwise captured in core demographic datasets play a role in driving COVID-19 racial inequities and that distinct analyses focused on systemic inequities are needed to fully understand the specific risks faced by Black individuals and communities.

Reduced risk of COVID-19 incidence in communities with higher percentages of Black residents, adjusting for other factors, may suggest that the clear racial disparity observed in early months of the pandemic has diminished over time in Massachusetts.<sup>5,24-26</sup> Our finding parallels other observations as to the reduced racial disparity in COVID-19 cases over

time.<sup>27</sup> Likewise, substantial reductions in the association between long-term care beds and percentage of town population over 80 years and COVID-19 incidence may reflect success in interventions to protect the elderly, especially those living in long-term or nursing care, after the initial, devastating impacts on this population early in the pandemic.<sup>28,29</sup> It remains possible that factors not included in our analysis, notably biases associated with testing availability or other unstudied correlates, are responsible for the reductions in risk we observe here, especially by race. Analysis of more severe outcomes, including hospitalization and deaths, would help inform these questions, although requiring less-nuanced temporal resolution, and should be prioritized for future work.

The continued positive association between essential workers and COVID-19 incidence may reflect the continued inability of this workforce to adequately protect themselves from infection, despite workplace controls and personal mitigation behaviors, such as masking and maintaining social distance. Essential workers remained at the highest risk of all subgroups studied in our models throughout the pandemic, highlighting key challenges in protecting these populations, even many months into the pandemic. Interestingly, however, a covariate representing people spending more than 3 hours in a location other than their home during office hours showed a null association with COVID-19 cases. The non-significant effect of worker mobility on case incidence may reflect limitations in the underlying data. SafeGraph data includes only 10% of cellphones in the US<sup>30</sup>, although the data are highly correlated with true Census population.<sup>31</sup> However, it could also indicate that changes in worker mobility (e.g. “return to work” efforts) were not associated with increased cases when these workers were not part of the essential workforce. This would reinforce that communities with more non-essential workers face distinctly lower risk than those with more essential workers, even with resumption of economic activities, highlighting inequities in exposure profiles on the job.

The significant association between the percentage of town residents without health insurance and case incidence in the spring and summer periods may suggest growing case incidence among immigrants during those times. Due to a state health insurance mandate, the number of non-insured individuals in Massachusetts is low (2.8% of the total state population), but some municipalities have uninsured rates up to 25%, such as Chelsea, Everett and Lawrence. These communities have higher proportions of younger adults, non-citizens, and those who are less educated than the insured population.<sup>32</sup> Targeted interventions that focus on communities with elevated percentages of uninsured persons are important in fully understanding disease dynamics in the state.

As noted, our study is limited by the use of town-level covariates, which reduce our ability to draw causal inferences on the individual level, though our insights remain relevant for targeted public health strategies. The use of static predictors from the ACS and other databases allowed us to evaluate and interpret changes in coefficient magnitude and significance over time, but gave us limited ability to capture time-dynamic exposure factors (with the exception of SafeGraph data). While most of these predictors are expected to be fairly stable during the study period, notably demographic data, it is possible that real-time changes in these covariates may have altered our findings. It is possible that disparities in testing availability, as noted earlier, contribute to observed trends in case incidence if communities with greater racial diversity had reduced access to case identification, and this is an important area for additional work.

## Conclusions

Our analyses demonstrate that the relevance and magnitude of risk factors on the community-level for COVID-19 are alterable, suggesting in part the relative effectiveness of intervention and mitigation efforts by population subgroups (or, conversely, factors that contribute to elevated risk varying over time). Our study highlights the need for local jurisdictions to use up-to-date data on vulnerable and high-risk populations to direct COVID-19 interventions, including vaccinations, rather than data from early in the pandemic. Our models were derived entirely from publicly-available data and could be rapidly refit to newer case incidence data. Community-level analyses can help characterize social inequities embedded in the pandemic and track the evolution of these inequities with time, highlighting successes as well as disproportionate burden experienced by vulnerable populations.

Funding contributions: This project is supported by NIH/NIMHD grant P50MD010428 (PIs: JIL and F. Laden) and by a gift from Google LLC. BH and PB were supported by NIEHS grant T32ES014562.

## Declarations

Funding contributions: This project is supported by NIH/NIMHD grant P50MD010428 (PIs: JIL and F. Laden) and by a gift from Google LLC. BH and PB were supported by NIEHS grant T32ES014562.

Data Availability: The datasets used during the current study are derived from publicly available data. Compiled datasets are available from the corresponding author on reasonable request

Competing Interests: The authors declare that they have no competing interests

Author contributions: KFT led and conducted data analyses, with support from PP, JIL, PB and JHL. FC, PB, BH and KS prepared datasets and guided their use in modeling. JIL, MPF, KJL and JHL conceptualized the project. JHL wrote the manuscript with participation from all authors.

Acknowledgements: The authors gratefully acknowledge colleagues at the Massachusetts Department of Public Health for their guidance and support of this analysis.

## References

1. Johns Hopkins Coronavirus Resource Center. COVID-19 Map. Johns Hopkins Coronavirus Resource Center. Accessed January 21, 2021. <https://coronavirus.jhu.edu/map.html>
2. Xu J, Hussain S, Lu G, et al. Associations of Stay-at-Home Order and Face-Masking Recommendation with Trends in Daily New Cases and Deaths of Laboratory-Confirmed COVID-19 in the United States. *Explor Res Hypothesis Med*. Published online July 8, 2020:1-10. doi:10.14218/ERHM.2020.00045
3. Lyu W, Wehby GL. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. *Health Affairs*. 2020;39(8):1419-1425. doi:10.1377/hlthaff.2020.00818
4. White ER, Hébert-Dufresne L. State-level variation of initial COVID-19 dynamics in the United States. *PLOS ONE*. 2020;15(10):e0240648. doi:10.1371/journal.pone.0240648
5. Wadhera RK, Wadhera P, Gaba P, et al. Variation in COVID-19 Hospitalizations and Deaths Across New York City Boroughs. *JAMA*. 2020;323(21):2192. doi:10.1001/jama.2020.7197
6. Massachusetts Department of Public Health. Weekly COVID-19 Public Health Report. Published January 28, 2021. <https://www.mass.gov/doc/weekly-covid-19-public-health-report-january-28-2021/download>
7. Figueroa JF, Wadhera RK, Lee D, Yeh RW, Sommers BD. Community-Level Factors Associated With Racial And Ethnic Disparities In COVID-19 Rates In Massachusetts. *Health Affairs*. 2020;39(11):1984-1992. doi:10.1377/hlthaff.2020.01040
8. Massachusetts Department of Public Health. COVID-19 Response Reporting. Accessed January 22, 2021. <https://www.mass.gov/info-details/covid-19-response-reporting>
9. US Census Bureau. 2014-2018 American Community Survey 5-year Data. Accessed December 22, 2020. [api.census.gov/data/2018/acs/acs5?](https://api.census.gov/data/2018/acs/acs5?)
10. Chambers L. Data Show COVID-19 Is Hitting Essential Workers and People of Color Harder, ACLU Massachusetts. Published 2020. Accessed November 13, 2020. [https://data.aclum.org/2020/04/07/covid-19-disproportionately-affects-vulnerable-populations-in-boston/?ms\\_aff=MA&initms\\_aff=MA&ms\\_chan=tw&initms\\_chan=tw](https://data.aclum.org/2020/04/07/covid-19-disproportionately-affects-vulnerable-populations-in-boston/?ms_aff=MA&initms_aff=MA&ms_chan=tw&initms_chan=tw)
11. MassGIS. Massachusetts Prison Locations. Published 11 2020. Accessed January 21, 2021. <https://docs.digital.mass.gov/dataset/massgis-data-prisons>

12. MassGIS. Long Term Care Residences. Published 05 2019. Accessed January 21, 2021. <https://docs.digital.mass.gov/dataset/massgis-data-long-term-care-residences>
13. Mass.gov. Information about facilities licensed or certified by the Division of Health Care Facility Licensure and Certification. Published February 2021. Accessed February 5, 2021. <https://www.mass.gov/service-details/find-information-about-licensed-or-certified-health-care-facilities>
14. SafeGraph. Social Distancing Metrics. Published 2020. Accessed November 15, 2020. <https://docs.safegraph.com/docs/social-distancing-metrics>
15. Kloog I, Chudnovsky AA, Just AC, et al. A New Hybrid Spatio-Temporal Model For Estimating Daily Multi-Year PM(2.5) Concentrations Across Northeastern USA Using High Resolution Aerosol Optical Depth Data. *Atmos Environ (1994)*. 2014;95:581-590. doi:10.1016/j.atmosenv.2014.07.014
16. US Census Bureau. 2010 Census - Block Maps. *The United States Census Bureau*. <https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-block-maps.html>
17. US Census Bureau. *City and Town Population Totals: 2010-2019*. <https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html>
18. Johnston R, Jones K, Manley D. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant*. 2018;52(4):1957-1976. doi:10.1007/s11135-017-0584-6
19. R Core Team. *R: The R Project for Statistical Computing*. <https://www.r-project.org/>
20. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *ArXiv e-prints*. 2014;arXiv:1406. doi:10.18637/jss.v067.i01
21. Tirupathi R, Muradova V, Shekhar R, Salim SA, Al-Tawfiq JA, Palabindala V. COVID-19 disparity among racial and ethnic minorities in the US: A cross sectional analysis. *Travel Medicine and Infectious Disease*. 2020;38:101904. doi:10.1016/j.tmaid.2020.101904
22. Lei H, Xu X, Xiao S, Wu X, Shu Y. Household transmission of COVID-19-a systematic review and meta-analysis. *J Infect*. Published online August 25, 2020. doi:10.1016/j.jinf.2020.08.033
23. Li W, Zhang B, Lu J, et al. Characteristics of Household Transmission of COVID-19. *Clin Infect Dis*. 2020;71(8):1943-1946. doi:10.1093/cid/ciaa450
24. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA*. 2020;323(24):2466-2467. doi:10.1001/jama.2020.8598
25. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891-1892. doi:10.1001/jama.2020.6548
26. Owen WF Jr, Carmona R, Pomeroy C. Failing Another National Stress Test on Health Disparities. *JAMA*. 2020;323(19):1905-1906. doi:10.1001/jama.2020.6547
27. The Atlantic. The COVID Tracking Project. Published February 4, 2020. Accessed February 4, 2021. <https://covidtracking.com/race/dashboard#state-ma>
28. McMichael TM, Currie DW, Clark S, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. *N Engl J Med*. 2020;382(21):2005-2011. doi:10.1056/NEJMoa2005412
29. Louie JK, Scott HM, DuBois A, et al. Lessons from Mass-Testing for COVID-19 in Long Term Care Facilities for the Elderly in San Francisco. *Clinical Infectious Diseases*. 2020;(ciaa1020). doi:10.1093/cid/ciaa1020
30. Google Colaboratory. Accessed January 15, 2021. [https://colab.research.google.com/drive/1u15afRytJMSizySFqA2EPIXSh3KTmNTQ#sandboxMode=true&scrollTo=\\_f8ToOpVuPD](https://colab.research.google.com/drive/1u15afRytJMSizySFqA2EPIXSh3KTmNTQ#sandboxMode=true&scrollTo=_f8ToOpVuPD)
31. What about bias in the SafeGraph dataset? Accessed January 21, 2021. <https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset>

## Figures

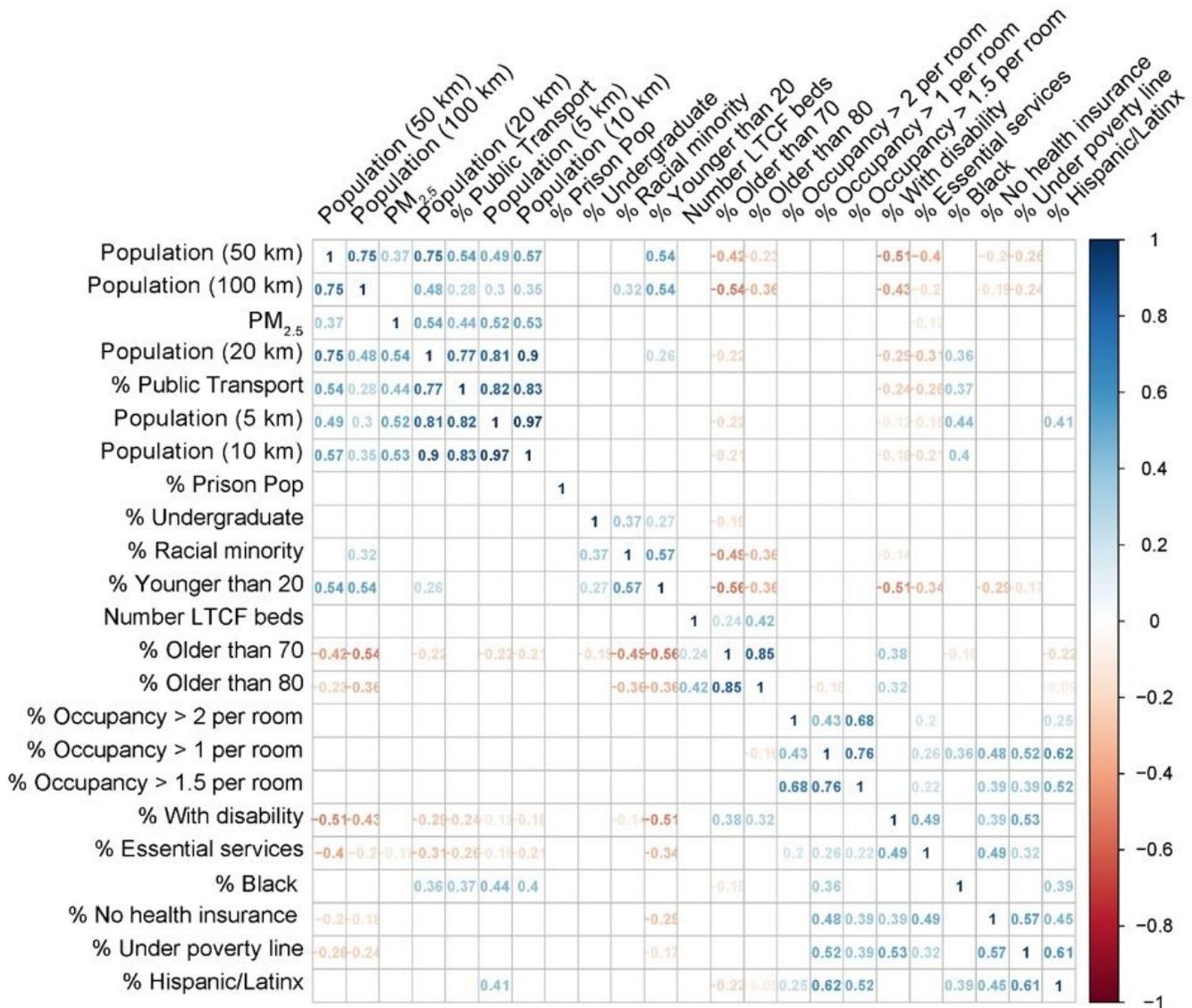
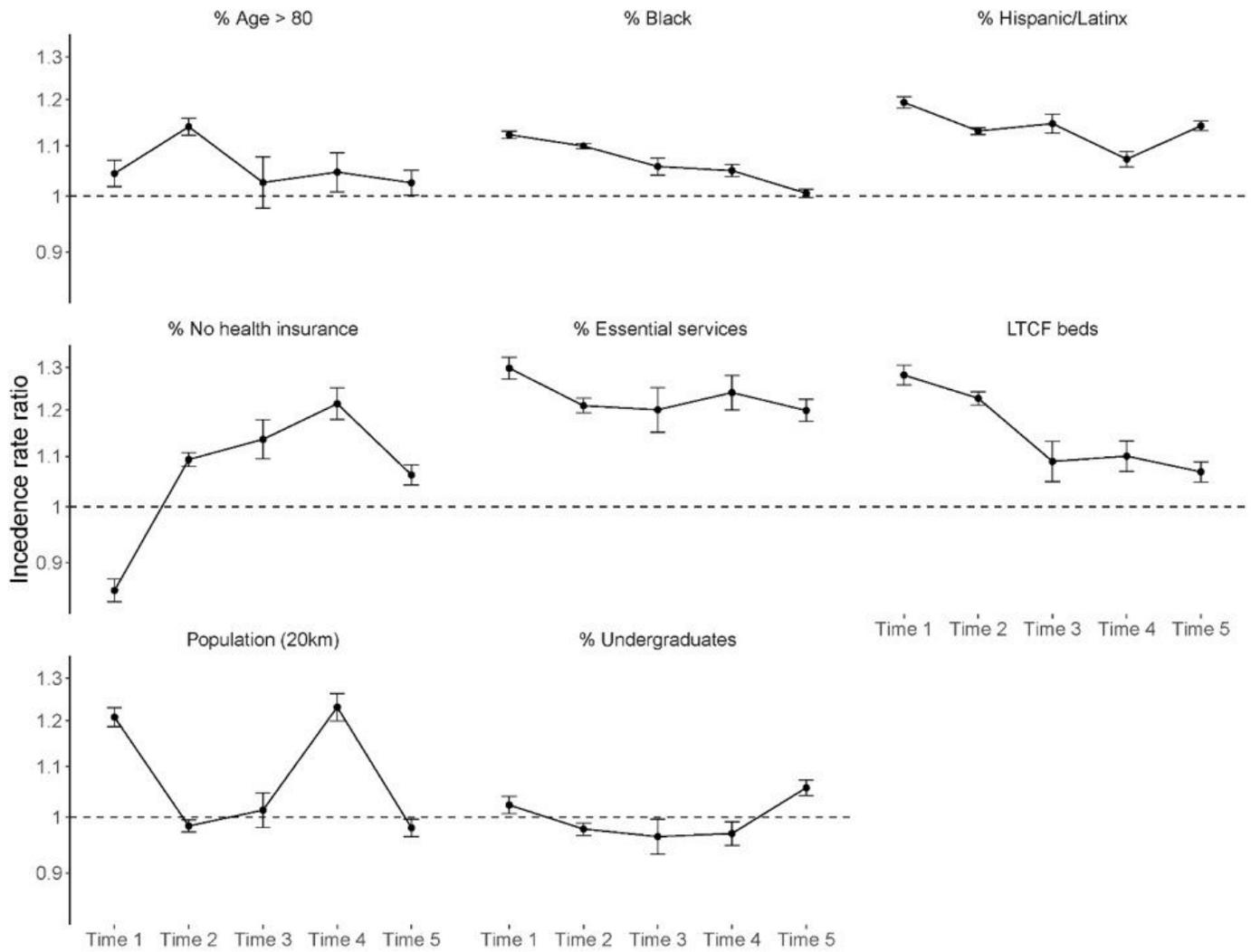


Figure 1

Pearson correlation coefficients for bivariate combinations of predictor variables (only showing correlations where  $p < 0.001$ )



**Figure 2**

Temporal change of incidence rate ratios (with 95% confidence intervals) of mixed effects models predicting the number of COVID-19 cases per town weighted by population during five phases of the pandemic in Massachusetts