Science of the Total Environment 811 (2022) 152284

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Investigating the association between mass transit adoption and COVID-19 infections in US metropolitan areas

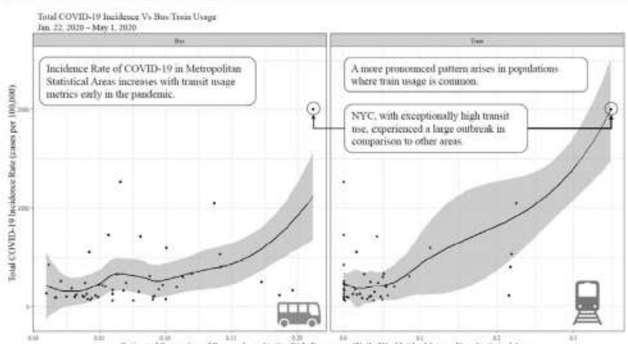
Michael M. Thomas, Neda Mohammadi, John E. Taylor *

School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr NW, Atlanta, GA 30332, United States

HIGHLIGHTS

- First national-scale US study to compare mass transit use and COVID-19.
- Mass transit use in US metropolitan areas associated with early COVID-19 incidence.
- Association remained valid with New York City metropolitan area data removed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 October 2021
Received in revised form 3 December 2021
Accepted 5 December 2021
Available online 10 December 2021

Editor: Jay Gan

Keywords:
SARS-CoV-2
COVID-19
Transit
Disease transmission
Public health
Public transportation

ABSTRACT

Urbanization introduces the threat of increased epidemic disease transmission resulting from crowding on mass transit. The coronavirus disease 2019 (COVID-19) pandemic, which has directly led to over 600,000 deaths in the US as of July 2021, triggered mass social distancing policies to be enacted as a key deterrent of widespread infections. Social distancing can be challenging in confined spaces required for transportation such as mass transit systems. Little is published regarding the degree to which mass transit system adoption effects impacted the rise of the COVID-19 pandemic in urban centers. Taking an ecological approach where areal data are the unit of observation, this national-scale study aims to measure the association between the adoption of mass transit and COVID-19 spread through confirmed cases in US metropolitan areas. National survey-based transit adoption measures are entered in negative binomial regression models to evaluate differences between areas. The model results demonstrate that mass transit adoption in US metropolitan areas was associated with the magnitude of outbreaks. Higher incidence of COVID-19 early in the pandemic was associated with survey results conveying higher transit use. Increasing weekly bus transit usage in metropolitan statistical areas by one scaled unit was associated with a 1.38 [95% CI: (1.25, 1.90)] times increase in incidence rate of COVID-19; a one scaled unit increase in weekly train transit usage was associated with an increase in incidence rate of 1.54 [95% CI: (1.42, 2.07)] times. These conclusions should inform early action practices in urban centers with busy transit systems in the event of future infectious disease outbreaks. Deeper understanding of these observed associations may also benefit modeling efforts by allowing researchers to include mathematical adjustments or better explain caveats to results when communicating with decision makers and the public in the crucial early stages of an epidemic.

* Corresponding author.
E-mail addresses: mthomas362@gatech.edu (M.M. Thomas), nedam@gatech.edu (N. Mohammadi), jet@gatech.edu (J.E. Taylor).

http://dx.doi.org/10.1016/j.scitotenv.2021.152284

1. Introduction

Diseases that spread through aerosols, droplets, or fomites are often transmitted widely in crowded environments. Therefore, social distancing has been a key feature of prevention strategies for coronavirus disease 2019 (COVID-19). Despite these efforts, as of July 2021 the COVID-19 pandemic directly led to over 600,000 deaths in the US according to Johns Hopkins Center for Systems Science and Engineering's continually updated dashboard (Dong et al., 2021). One important aspect of outbreak prevention involves targeting locations with increased transmission rates with information, testing, and other preventative or assessment measures. Although transmission rates of novel diseases are generally estimated on a population level, an emerging epidemic of such a disease can be challenging to characterize because the key analytic parameters estimated by early studies are specific to the initial outbreak's population. If ungeneralizable data are used to model transmission rates, it is possible that inadequate resources will be allocated to locations with greater disease burden (Jewell et al., 2020).

Early estimates of COVID-19 transmission rates were based on data from Wuhan, a city with a popular mass transit system (BBC, 2020; Yang et al., 2014). As soon as the intensity of the local epidemic was acknowledged, transit systems in many major cities were reduced or temporarily shut down to avoid increased disease spread. Once disease spread was marginally contained, authorities requested improved ventilation, sanitation, and social distancing where possible (BBC, 2020; Calvert, 2020; Chan et al., 2020; Shen et al., 2020). Although similar preventative measures were taken in many urban centers, COVID-19 was prevalent in many communities before political and public health agencies acted (Carteni et al., 2020; Jorden et al., 2020; Teixeira and Lopes, 2020). The spread of COVID-19 on transit can be especially challenging to control due to rapid spread and asymptomatic carriers who are not aware of their need to self-quarantine (Chipimo et al., 2020). It became clear as the pandemic unfolded that transmission rates do vary from population to population, so including transit use among other factors including race, age, essential worker status, and household structure is integral to estimating where disease prevention and medical resources are needed (Dasgupta et al., 2020; Grijalva et al., 2020; Lewis et al., 2020; Oster, 2020; Scarpone et al., 2020; Stephanie et al., 2020).

Researchers have investigated the impact of transit acquired infections on previous outbreaks and air transmitted endemic diseases such as tuberculosis in the literature (Andrews et al., 2013; Zamudio et al., 2015). Because of challenges involved in contact-tracing large crowds, there have been few studies that are able to directly estimate the risks of mass transit use in high disease prevalence areas. Moser and colleagues' work demonstrated a nearly indisputable instance of a single infectious individual spreading influenza, another disease that transmits well in close proximity, to a number of other riders on an unventilated aircraft (Moser et al., 1979). The circumstances surrounding this study are rare; however, the various facets of non-COVID-19 transit disease transmission risks have been illuminated using myriad alternative methods.

Numerous epidemiological study designs have been used to avoid the challenges posed by contact tracing large crowds to estimate transit adoption's effect on disease spread. Traditional mathematical modeling includes population dynamics methods to estimate the impact of travel on epidemics in ways that avoid having to collect specific individual-level ridership data (Xu et al., 2013). To further address issues of defining ridership habits of study participants and add granularity to the mathematical approach, simulation studies explained the impact of transit adoption in major cities with rail systems using agent-based models that leveraged contact probabilities to estimate spread (Cooley et al., 2011; Yashima and Sasaki, 2014). Similar computational methods in combination with household surveys led to simulation-based contact estimation; these studies found that mass transit acquired infections made up a small but non-negligible proportion of simulated disease spread in the UK and South Africa respectively (Johnstone-Robertson et al., 2011; Mossong et al., 2008). Cohort and case-control studies have been executed to draw contrast between transmissible disease incidence rates among populations who travel in close proximity to others and populations that do not (Baker et al., 2010; Horna-Campos et al., 2007; Hu et al., 2020). Each of these analyses suggests a relationship between infectious disease and transit adoption, but none focus on COVID-19.

Although there was little published on COVID-19 and transit adoption, researchers conducted literature reviews to summarize findings that convey transportation disease transmission risks or best practices to public health officials (Browne et al., 2016; Edelson and Phypers, 2011; Mohr et al., 2012; Nasir et al., 2016; Tirachini and Cats, 2020). Many of these studies inferred that crowded modes of transit, poor ventilation, and long travel times each increases disease transmission on mass transit; however, there are few that have examined this association in the context of COVID-19. Ecological studies have examined the hypothesis that there is an association between transit adoption and disease transmission by comparing measures of disease between geographic units, some granular enough to define proximity to a rail station and exposure variable (Sanna and Hsieh, 2017); however, this ecological study was not specific to COVID-19. The association between COVID-19 infections and mass transportation has been established in China (Zheng et al., 2020). But further national-scale study that considers covariates via explicit modeling of control variables has yet to be published. Implementing an ecological study allows for quick, large scale analysis of emerging diseases that the aforementioned research methods cannot obtain. However, measuring exposure to transit in an ecological analysis can also be a challenge given there is little available data directly measuring the entire populations of riders' mass transit adoption.

The literature on transit modes as an exposure in ecological studies often relies on survey data to quantify exposure to different modes of transit, because those data offer a level of detail unavailable in most administrative data (e.g., ridership estimates by a transit authority). This is relevant to disease modeling because using a diverse array of exposure estimation methods limits potential biases in assessing exposure-outcome relationships. For example, there are numerous studies that examined the link between ride time and location to bicycle accident frequency, exploiting time spent on transportation mode as a control variable (Beck et al., 2007; Schneider et al., 2017). Weighted surveys including the National Household Travel Survey (NHTS) have been leveraged to estimate similar risk factors such as ride distance where a simple random sample is infeasible (Buehler and Pucher, 2017; Pucher and Renne, 2003). Researchers also strengthened their transit-related hypotheses by using multiple exposure measures to describe similar constructs (Ferenchak and Marshall, 2020). Many of these analyses are similar to ecological analyses of COVID-19 spread and transit in that they measure the relationship between transit mode adoption and some health outcome. Researchers have used ridership estimates to determine whether the number of transit riders increasing led to sharp increases in COVID-19 cases, but little effect was found after controlling for covariates such as essential worker density (Sy et al., 2020). Other studies have considered urban factors, but did not focus on mass transportation (Hu et al., 2021; Scarpone et al., 2020). This study leverages survey-based exposure measures as opposed to ridership estimates in a national-scale ecological analysis to address this gap in the literature and account for this finding.

Understanding factors that impact the disease spread helps modelers create more accurate estimates, and therefore, will improve the accuracy of information used by policymakers. Mass transit adoption may be an important factor in estimating disease spread parameters. Although transit has been considered a potential risk factor for the spread of COVID-19, there is currently a gap in the research where this association has not been shown for COVID-19 specifically on a scale larger than one city or subnational region using exposure data regarding transit adoption attitudes. This study applies statistical modeling to assess the association between transit adoption and COVID-19 incidence while addressing potential confounding variables that may affect the association on a metropolitan-area level. A hierarchical regression model compares COVID-19 incidence early in the pandemic using transit attitude measures for metropolitan statistical areas (MSA) in the US.

2. Material and methods

2.1. Data

The data used to model the association between mass transit adoption and COVID-19 incidence rates were gathered from the NHTS and the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) (Dong et al., 2021; U.S. Dept of Transportation, 2017). Between the beginning of JHU's data collection period, January 22, 2020, and May 1, 2020, a total of 837,811 cases were recorded in 52 MSAs; these values are aggregated into county-based bins which prevents individual-level analysis. These NHTS includes responses to the survey item "Frequency of ___ Use for Travel" for 52 of 384 MSAs. The question asked individuals how often they rode the train or bus on a Likert-type scale with options like "Daily," "A few times a week," "A few times a month," "A few times a year," and "Never." "Daily" and "A few times a week" were grouped in this study, because frequent public transportation use is relatively uncommon in many U.S. cities (Blumenberg et al., 2020), and a combined group retains the same direction increase in transit usage behavior.

MSAs are combinations of counties created by the U.S. Census Bureau to represent urbanized areas containing 50,000 or more residents (U.S. Census Bureau, 2020). The early time period was chosen to demonstrate effects that included disease transmission before and during individuals' changing attitudes towards transit and policies preventing them from using transit. Because many of the public health interventions and changes in human behavior happened early in the pandemic, it is likely that this exposure type did occur more in the first few months of the pandemic than later when policy and risks were clearer (Mohammadi and Taylor, 2020). Data from the NHTS was estimated using the *summarizeNHTS* R package (Fucci and Cates, 2018) to express transit habits of people living in the MSAs; the most recent version of the survey available, the 2017 version, including responses from 129,626 individuals which were weighted by an algorithm aiming to compensate for their sampling probability and thereby improve the representativeness of the responses (U.S. Dept of Transportation, 2017). The most recent version of the travel surveys are often used to measure travel outcomes even years after their publication (Blumenberg et al., 2020; Pucher and Buehler, 2016). The proportions of households answering that they ride a bus weekly or daily were summed to measure the proportion of riders among the population of each MSA shown in Figs. 1 and 2. The same transformation was applied to the NHTS train/light rail question.

Data from the 2019 American Community Survey 5-year estimates, the most recent complete U.S. Census data at the time of this study's completion, were aggregated at the MSA level and attached to the dataset as features and used in the modeling process to control for other COVID-19 risk factors on an MSA scale (U.S. Census Bureau, 2019). The first covariate, percentage of families in the last 12 months that are below the poverty level was selected to control for poverty (Sy et al., 2020). Percentage of individuals over 25 with only a high school diploma (including equivalency

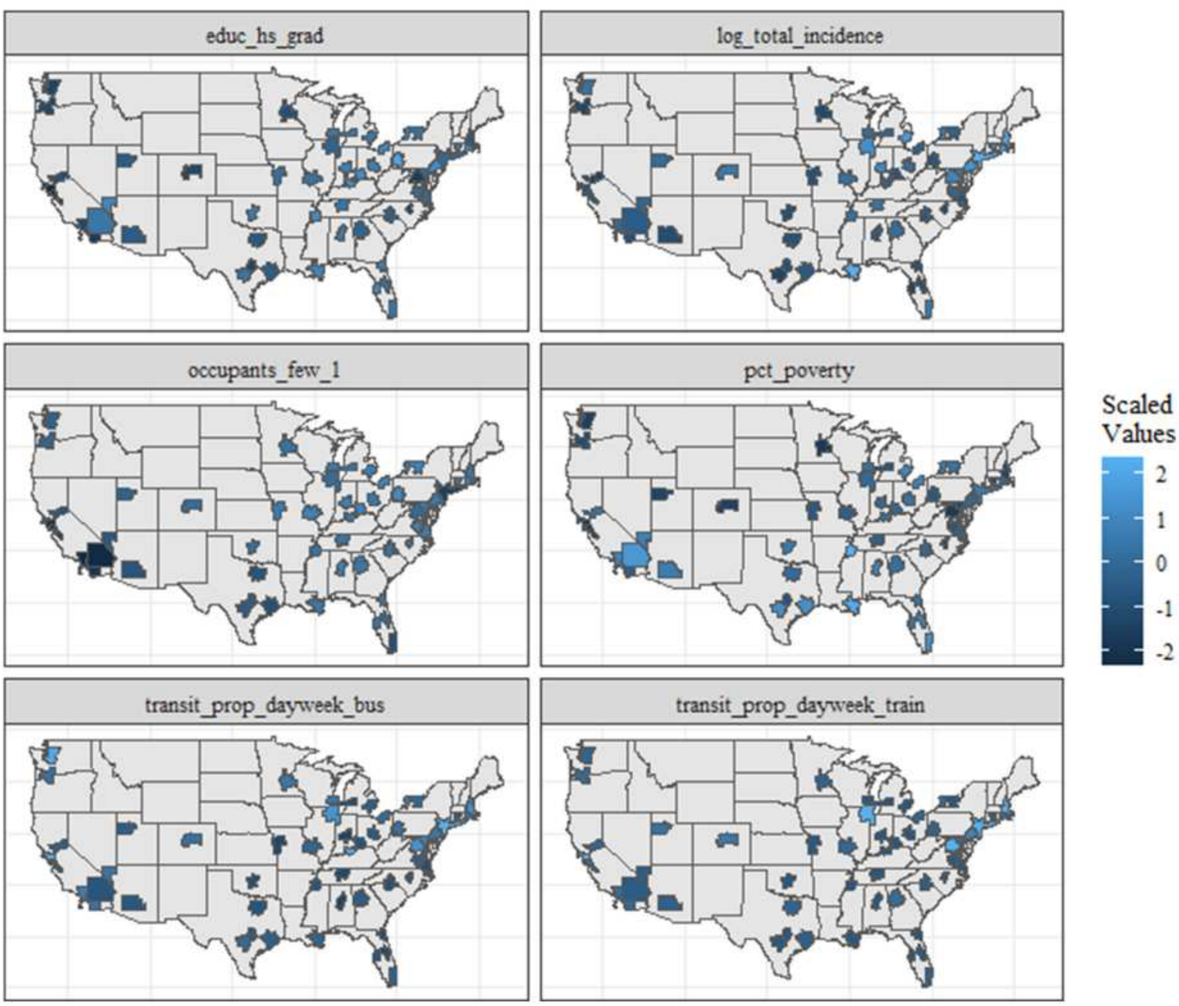


Fig. 1. Spatial Distribution of Scaled Values (left-to-right starting from top left) percentage of individuals over 25 with only a high school diploma (including equivalency exam), COVID-19 incidence rate, percentage of households with one occupant or fewer per room, percentage of families in the last 12 months that are below the poverty level, percentage of survey respondents stating transit (Bus/Train) use daily or weekly. All values are mean normalized.

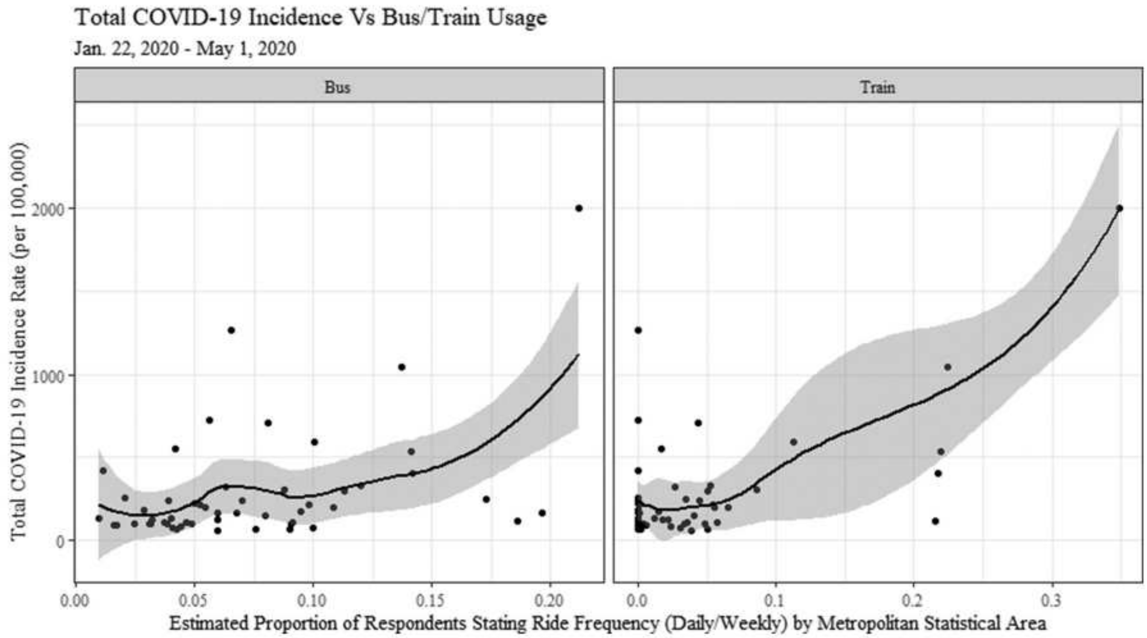


Fig. 2. Total COVID-19 incidence plotted against bus and train usage with LOESS smoothed line with standard error estimate describing association.

exam) was selected as a measure of education as a proxy for essential worker status (Yang et al., 2021). Percentage of households with one occupant or fewer per room was selected to represent density of residential contact networks (Mossong et al., 2008; Rojas-Bolivar et al., 2021). Covariates were selected based on presence in the literature; the number of covariates was limited to the three cited in ecological studies above to prevent multicollinearity and model convergence issues that occur with small sample sizes (Rojas-Bolivar et al., 2021; Yang et al., 2021).

2.2. Analysis methods

First, a locally estimated regression method known as LOESS lines and standard errors were applied to a plot of incidence rate versus the proportion of respondents stating they take mass transit either weekly or daily (Jacoby, 2000). LOESS smoothing is a non-parametric data visualization method used to describe the relationship between two continuous variables. A set of equally distant points v_j are defined on the range of an input variable, X, then a linear regression model is trained at each of the j points using the values of X and the response variable, Y. Only values that are "close" to the point being estimated are included in the regression at point j. The parameter that determines what is "close to" j is called the span and is represented by the Greek character, α. The span represents the proportion of the data points used in each of the j regression models. The parameter α for this study was selected by generating plots with values (0.45, 0.65, 0.75, 0.95). The 0.75 plot was the first to remove the majority of small variations in the fitted line among the cluster points on the left side of the plot which were likely results of overfitting.

Negative binomial regression was used to estimate the association between mass transit adoption and COVID-19 incidence. The negative binomial regression model was chosen for its common use in ecological studies of disease involving incidence rates for diseases in areal data as well as its ability to replace the Poisson regression model when response data are overdispersed (Liu et al., 2020; Szklo and Nieto, 2014). For the negative binomial regression models, MSA-level estimates of COVID-19 incidence in US MSAs are modeled as the response variable $Y_i \sim log(\mu_i)$ with an offset term for the population of each MSA. A negative binomial model is a generalized linear model with a log link and response modeled as a binomial random variable. Random effects error modeling is used to address the regression assumption violation of independent observations. In the context of this analysis, that means that the random effect term will be used to model the spatial autocorrelation attributable to differences in U.S. Census Regions that are not explicitly indicated among the other independent predictors such as climate or elevation. In this model, case counts in each MSA, Y_i, are assumed to be independent random variables each with mean μ_i. The model is stated as

$$log(Y_i) = \beta_{intercept} + \beta_{transit} X_{i,transit} + \sum_{j=1}^{m} \beta_{cov_j} X_{cov_{ij}} + b_{i,censusReg} + O(n_i).$$

In the model, each of the unknown parameters representing fixed effects is represented by β. $\beta_{transit}$ measures the association between mass transit usage (i.e., bus or train), β_{covj} measures each of the m covariates' impact on cases, $bi_{,censusReg}$ controls for the variance attributable to Census Region, and an offset term $O(n_i)$ is included to control for the size of an area's population (n_i). The offset term is used to define the output in terms of impact on the *rate* for each location as opposed to raw case counts; this is necessary because the population in an area impacts the number of cases that arise. The primary exposure of interest is proportion of survey respondents stating weekly or daily bus or train transit usage X_{bus}, X_{train} (i.e., they responded "daily" or "weekly" to the survey item asking about their transit use frequency). The negative binomial regression model parameters were estimated using iteratively reweighted least squares. Base R and *lme4* software packages were used to estimate the parameters for the model (Bates et al., 2015; R Core Team, 2018).

3. Results

The maps in Fig. 1 show that there are large disparities in mass transit use behavior across the 52 MSAs. The usage difference is exaggerated in the train usage plot. A relatively small number of MSAs have greater than 20% of frequent riders. The figure drops down to approximately 10% for Philadelphia-Camden-Wilmington.

Fig. 2, the LOESS smoothed regression line, plots the bivariate association between train and bus rider attitudes in MSAs and COVID-19 spread early in the pandemic. The curve, coinciding with common population dynamics and epidemiologic patterns, follows an exponential increase in incidence rates resulting from both mass transit settings (Szklo and Nieto, 2014). Both LOESS smoothed lines are heavily influenced by cities with high transit usage, especially New York City, which has especially crowded transit during busy times. This visual statistical output does have a relatively weaker but still positive association in the smaller proportion values, but once about 10% of the respondents indicate they ride mass transit weekly or daily, the association is evident.

The train mass transit adoption and bus mass transit adoption negative binomial regression models demonstrate a positive association between total COVID-19 incidence and mass transit adoption through their corresponding model parameters (Table 1). Both terms, $\widehat{\beta_{bus}} = 0.43$ [95% CI: (0.23, 0.64)] and $\widehat{\beta_{train}} = 0.53$ [95% CI: (0.35, 0.73)] are statistically significant and relatively high magnitude implying that incidence early in the pandemic did increase with transit usage. The 0.43 value implies that increasing the bus transit weekly or daily riders by one scaled unit increases the estimated incidence rate by 1.38 [95% CI: (1.25, 1.90)] times. The train estimate, 0.53, implies that a one scaled unit increase in train transit weekly or daily riders increases the estimated incidence rate by 1.54 [95% CI: (1.42, 2.07)] times. This result adds to the result of the LOESS smoothed lines, because it also takes into account other known risk factors of large COVID-19 outbreaks in urban centers including household crowding, poverty, and education. Despite controlling for potential confounding factors, there was still a clear association between transit adoption and COVID-19 incidence early in the pandemic. Because of the apparent influence of the New York City MSA data point in the LOESS smoothed plots, the model was run without the data point for comparison; the statistically significant association remained regardless of the observation's presence in the model (See table in Appendix I).

4. Discussion

This study represents the first national-scale assessment of the association between mass transit adoption in MSAs and COVID-19 incidence rate in the US controlling for the effects of other known risk factors. A measure of COVID-19 incidence rate was regressed against mass transit adoption measures to establish this connection. Studies within individual regions have proposed similar results and used transit attitude data to investigate transit outcomes on a metropolitan area scale (Buehler and Pucher, 2017; Sy et al., 2020). This study built upon a combination of these methods to evaluate a similar hypothesis on a national scale across the MSAs. The research of Sy et al. (2020) focused on the impact of transit adoption for disadvantaged communities on COVID-19 outcomes. The current study applied a methodology examining a similar association on a national scale; similar to Buehler and Pucher's (2017) work, this study leveraged the NHTS to examine the impact of transit adoption on an outcome. Buehler and Pucher acknowledge in their work that there is benefit in examining exposure data that inspects specific aspects of individual level transit such as the number of trips per household per week, which was used in this paper.

LOESS smoothed visualizations and negative binomial regression model results were congruent in supporting the hypothesis that mass transit adoption in US MSAs was positively associated with COVID-19 incidence early in the pandemic. The train transit and bus transit LOESS plots do demonstrate an inflection point that shows an increase in growth including the MSAs of New York City, Chicago, Washington, and Boston. Although it is clear that the transit adoption in New York City MSA and high COVID-19 incidence influences the magnitude of this association in this study, this MSA represents an extreme case of mass transit adoption. That is, based on the survey, transit crowding is likely highest in New York City MSA, and therefore, that municipal area would suffer the greatest disease spread given that the hypothesis of this study is correct.

Table 1
Negative binomial regression model results.

	Estimate	SE	95% CI	*P*-Value
Train Negative Binomial ($N = 52$)				
(Intercept)	−6.06	0.10	(−6.30, −5.84)	<0.001
Proportion Ride Train Daily or Weekly	0.53	0.09	(0.35, 0.73)	<0.001
Proportion Below Poverty Line	0.23	0.12	(0.00, 0.47)	0.058
Proportion with <1 Occupant per Room	0.22	0.12	(−0.02, 0.46)	0.071
Proportion High School Graduate	0.12	0.17	(−0.21, 0.45)	0.462
Bus Negative Binomial (N = 52)				
(Intercept)	−6.03	0.15	(−6.37, −5.68)	<0.001
Proportion Ride Bus Daily or Weekly	0.43	0.11	(0.23, 0.64)	<0.001
Proportion Below Poverty Line	0.21	0.14	(−0.06, 0.49)	0.139
Proportion with <1 Occupant per Room	0.11	0.15	(−0.19, 0.39)	0.111
Proportion High School Graduate	0.02	0.20	(−0.39, 0.41)	0.016

The results of this study are of value to epidemiologists, policymakers, and municipal transit providers. Modeling attempts made earlier in the pandemic were integral to the allocation of resources in early stages of transmission in the US. Literature has criticized some of these models for using data from southern Italy and Hubei province to inform large areas that differ in key factors relating to the application of social distancing (Jewell et al., 2020). Paying attention to these factors, such as mass transit adoption in a region, when generating or presenting model results may improve the way policymakers are able to avoid mistakenly sending resources to areas with fewer transmission risks leaving areas with high transmission risk without aid. Epidemiologic modelers can modify disease transmission parameters with mathematical adjustments or clarify rhetorically the impacts of the generalizability when presenting models to policy makers and the public. Additionally, it may be possible to use the NHTS estimates of transit to weight the disease transmission rates in models explicitly outlining their impact. Awareness regarding the risks of transit can keep the aforementioned entities ahead of heuristic assessments of the risks of riding transit. Preempting these risks could help prevent the magnitude of transit ridership drop-offs that occurred at the beginning of the pandemic and continued through the ensuing year (Sharifi and Khavarian-Garmsir, 2020). Some research has shown that these changes may persist, so it is imperative that those advocating for transit adoption remain aware of the risks and actively address them when they appear (Mohammadi and Taylor, 2020).

Although this study identifies mass transit adoption as one of many risk factors for COVID-19 transmission in cities, it should be mentioned that there are key steps that can curtail this effect. Tirachini and Cats (2020) as well as Nasir et al. (2016) each wrote review papers on changes mass transportation agencies can implement in the early stages of an epidemic or pandemic of a disease like COVID-19. Physical distancing of riders, sanitation of train cars/buses, improved ventilation, and mask usage for riders were each identified as important protective measures for transportation systems. Of course, each of these factors does involve a cost to the riders and operators of transit systems. As the authors of the aforementioned studies suggest, future research should be conducted to evaluate the efficacy of these interventions in light of the inherent tradeoffs attributable to each.

One limitation of the methods used in this study is a fluctuation in testing availability in the US early in the pandemic. Testing availability stabilized in the latter half of the study period where the majority of cases occurred, but the beginning of the pandemic in the US (early February and March) was characterized by low testing supply (Hadaya et al., 2020). Because the highest COVID-19 incidence rates occurred late in the study period, it is likely that this had little effect on the estimates for the entire period. Association in a geographic scale statistical analysis does not necessarily imply a causal link between a risk factor and an outcome. Despite taking measures to reduce confounding and finding a result that is consistent with the literature, there is still the possibility that this association is a function of a different variable regarding the MSAs. Due to the small number of MSAs, the models could only permit a small number of covariates; further study with a larger sample would benefit further research by increasing the number of potential confounding relationships that may modify the association. A few examples of potential confounding variables or effect modifiers include tourism in cities with better public transportation systems, demographic composition of cities, better healthcare access in more developed cities, and presence of diseases increasing COVID-19 risk. Because of this limitation, further research should take a more individual-focused approach to interrogating this potential link. An individual-level study would also allow for taking into account the distance and time of travel. A cohort study like that of Horna-Campos and colleagues where a number of transit riders and non-transit riders are linked to COVID-19 testing data may help better ensure the veracity of this result, however, there are myriad challenges in acquiring health data quickly and recruiting subjects (Horna-Campos et al., 2007).

5. Conclusion

The analysis in this study elucidated the association between the stated adoption of mass transit and COVID-19 incidence in U.S. metropolitan statistical areas; there is evidence that urban centers with high usage transit systems experience differential disease transmission rates than those without. This finding adds to existing findings by focusing specifically on COVID-19 as opposed to other infectious diseases, expanding the analysis to a national scale while controlling for relevant crowding and socioeconomic indicators. This notion implies that transit infrastructure must respond quickly when a potential pandemic situation is developing. Ventilation, social distancing, and other public health measures support meeting this need. This finding also suggests that modeling efforts undertaken to estimate the number of cases and, by extension, impact of policies early in a pandemic, such as the 2020 COVID-19 pandemic, should take measures to control for the impact of mass transit and possibly other urban crowding measures in their estimates. This directly relates to models generated for COVID-19, some of which used transmission rates from Wuhan, China, a region with higher transit usage than most US urban centers. Although their work was integral to the COVID-19 response, in order to retain the trust of the public and lawmakers, it is essential that epidemiologic modelers provide as accurate and consistent models as possible. Considering transit and other relevant population characteristics provides an essential component of meeting that end and ultimately reducing the burden of devastating health emergencies.

Credit authorship contribution statement

Michael M. Thomas: Conceptualization, Methodology, Data curation, Writing – original draft, Visualization, Investigation. **Neda Mohammadi:** Conceptualization, Methodology, Supervision, Writing – review & editing. **John E. Taylor:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding: This material is based upon work supported by the National Science Foundation under Grant No. 1837021. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix I

Table A1
Negative binomial regression model results (Without New York).

	Estimate	SE	95% CI	P-Value
	Train Negative Binomial ($N = 52$)			
(Intercept)	−6.07	0.11	(−6.32, −5.84)	<0.001
Proportion Ride Train Daily or Weekly	0.50	0.13	(0.26, 0.75)	<0.001
Proportion Below Poverty Line	0.23	0.12	(−0.01, 0.48)	0.063
Proportion with <1 Occupant per Room	0.23	0.13	(−0.02, 0.48)	0.066
Proportion High School Graduate	0.11	0.17	(−0.24, 0.44)	0.539
	Bus Negative Binomial (N = 52)			
(Intercept)	−6.08	0.14	(−6.40, −5.76)	<0.001
Proportion Ride Bus Daily or Weekly	0.33	0.11	(0.12, 0.55)	0.003
Proportion Below Poverty Line	0.20	0.13	(−0.05, 0.47)	0.130
Proportion with <1 Occupant per Room	0.21	0.15	(−0.09, 0.49)	0.145
Proportion High School Graduate	−0.03	0.19	(−0.40, 0.34)	0.891

References

Andrews, J.R., Morrow, C., Wood, R., 2013. Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. Am. J. Epidemiol. 177 (6), 556–561.

Baker, M.G., Thornley, C.N., Mills, C., Roberts, S., Perera, S., Peters, J., Kelso, A., Barr, I., Wilson, N., 2010. Transmission of pandemic A/H1N1 2009 influenza on passenger aircraft: retrospective cohort study. Br. Med. J. 340, c2424.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using {lme4}. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

BBC, 2020. Coronavirus: Wuhan shuts public transport over outbreak. BBC. https://www.bbc.com/news/world-asia-china-51215348.

Beck, L.F., Dellinger, A.M., O'Neil, M.E., 2007. Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. Am. J. Epidemiol. 166 (2), 212–218. https://doi.org/10.1093/aje/kwm064.

Blumenberg, E., Brown, A., Schouten, A., 2020. Car-deficit households: determinants and implications for household travel in the US. Transportation 47 (3), 1103–1125.

Browne, A., St-Onge Ahmad, S., Beck, C.R., Nguyen-Van-Tam, J.S., 2016. The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: a systematic review. J. Travel Med. 23 (1).

Buehler, R., Pucher, J., 2017. Trends in walking and cycling safety: recent evidence from high-income countries, with a focus on the United States and Germany. Am. J. Public Health 107 (2), 281–287. https://doi.org/10.2105/AJPH.2016.303546.

Calvert, S., 2020. Public transit agencies slash services, staff as coronavirus keeps ridership low. November 28Wallstreet J.. https://www.wsj.com/articles/public-transit-agencies-slash-services-staff-as-coronavirus-keeps-ridership-low-11606582853?mod=hp_lead_pos8.

Carteni, A., Di Francesco, L., Martino, M., 2020. How mobility habits influenced the spread of the COVID-19 pandemic: results from the italian case study. Sci. Total Environ. 741, 140489.

Chan, H.F., Skali, A., Savage, D., Stadelmann, D., Torgler, B., 2020. Risk Attitudes and Human Mobility during the COVID-19 Pandemic. 10. Scientific Reports - Nature Publishing Group, pp. 1–13.

Chipimo, P.J., Barradas, D.T., Kayeyi, N., Zulu, P.M., 2020. First 100 persons with COVID-19 - Zambia, march 18-april 28, 2020. MMWR Morb. Mortal. Wkly Rep. 69 (42), 1547–1548. https://doi.org/10.15585/mmwr.mm6942a5.

Cooley, P., Brown, S., Cajka, J., Chasteen, B., Ganapathi, L., Grefenstette, J., Hollingsworth, C.R., Lee, B.Y., Levine, B., Wheaton, W.D., et al., 2011. The role of subway travel in an influenza epidemic: a New York City simulation. J. Urban Health 88 (5), 982.

Dasgupta, S., Bowen, V.B., Leidner, A., Fletcher, K., Musial, T., Rose, C., 2020. Association between social vulnerability and a County's risk for becoming a COVID-19. Hotspot 69 (42), 1535–1541.

Dong, E., Du, H., Gardner, L., 2021. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20 (5), 533–534.

Edelson, P.J., Phypers, M., 2011. TB transmission on public transportation: a review of published studies and recommendations for contact tracing. Travel Med. Infect. Dis. 9 (1), 27–31.

Ferenchak, N.N., Marshall, W.E., 2020. Is bicycling getting safer? Bicycle fatality rates (1985–2017) using four exposure metrics. Transp. Res. Interdiscip. Perspect. 8, 100219.

Fucci, A., Cates, A., 2018. summarizeNHTS. GitHub. https://github.com/Westat-Transportation/summarizeNHTS.

Grijalva, C.G., Rolfes, M.A., Zhu, Y., Mclean, H.Q., Hanson, K.E., Belongia, E.A., 2020. Transmission of SARS-COV-2 Infections in Households — Tennessee and Wisconsin, April – September 2020. 69(44), pp. 1631–1634.

Hadaya, J., Schumm, M., Livingston, E.H., 2020. Testing individuals for coronavirus disease 2019 (COVID-19). J. Am. Med. Assoc. 323 (19), 1981.

Horna-Campos, O.J., Sánchez-Pérez, H.J., Sánchez, I., Bedoya, A., Martin, M., 2007. Public transportation and pulmonary tuberculosis, Lima, Peru. Emerg. Infect. Dis. 13 (10), 1491.

Hu, M., Lin, H., Wang, J., Xu, C., Tatem, A.J., Meng, B., Zhang, X., Liu, Y., Wang, P., Wu, G., Xie, H., Lai, S., 2020. Risk of coronavirus disease 2019 transmission in train passengers: an epidemiological and modeling study. Clin. Infect. Dis. https://doi.org/10.1093/cid/ciaa1057.

Hu, M., Roberts, J.D., Azevedo, G.P., Milner, D., 2021. The role of built and social environmental factors in Covid-19 transmission: a look at America's capital city. Sustain. Cities Soc. 65, 102580.

Jacoby, W.G., 2000. Loess: a nonparametric, graphical tool for depicting relationships between variables. Elect. Stud. 19 (4), 577–613.

Jewell, N.P., Lewnard, J.A., Jewell, B.L., 2020. Caution warranted: using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. Am. Coll. Physicians 173, 226–227.

Johnstone-Robertson, S.P., Mark, D., Morrow, C., Middelkoop, K., Chiswell, M., Aquino, L.D.H., Bekker, L.-G., Wood, R., 2011. Social mixing patterns within a south african township community: implications for respiratory disease transmission and control. Am. J. Epidemiol. 174 (11), 1246–1255.

Jorden, M.A., Rudman, S.L., Villarino, E., Hoferka, S., Patel, M.T., Bemis, K., Simmons, C.R., Jespersen, M., Johnson, J.I., Mytty, E., 2020. Evidence for Limited Early Spread of COVID-19 Within the United States, January–February 2020.

Lewis, N.M., Friedrichs, M., Wagstaff, S., Sage, K., LaCross, N., Bui, D., McCaffrey, K., Barbeau, B., George, A., Rose, C., Willardson, S., Carter, A., Smoot, C., Nakashima, A., Dunn, A., 2020. Disparities in COVID-19 incidence, hospitalizations, and testing, by area-level deprivation - Utah, March 3-July 9, 2020. MMWR Morb. Mortal. Wkly Rep. 69 (38), 1369–1373. https://doi.org/10.15585/mmwr.mm6938a4.

Liu, J., Zhou, J., Yao, J., Zhang, X., Li, L., Xu, X., He, X., Wang, B., Fu, S., Niu, T., et al., 2020. Impact of meteorological factors on the COVID-19 transmission: a multi-city study in China. Sci. Total Environ. 726, 138513.

Mohammadi, N., Taylor, J.E., 2020. Human-Infrastructure Interactional Dynamics: Simulating COVID-19 Pandemic Regime Shifts. 2020 Winter Simulation Conference (WSC), pp. 727–735.

Mohr, O., Askar, M., Schink, S., Eckmanns, T., Krause, G., Poggensee, G., 2012. Evidence for airborne infectious disease transmission in public ground transport–a literature review. Eurosurveillance 17 (35), 20255.

Moser, M.R., Bender, T.R., Margolis, H.S., Noble, G.R., Kendal, A.P., Ritter, D.G., 1979. An outbreak of influenza aboard a commercial airliner. Am. J. Epidemiol. 110 (1), 1–6.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., et al., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5 (3), e74.

Nasir, Z.A., Campos, L.C., Christie, N., Colbeck, I., 2016. Airborne biological hazards and urban transport infrastructure: current challenges and future directions. Environ. Sci. Pollut. Res. 23 (15), 15757–15766.

Oster, A.M., 2020. Transmission dynamics by age group in COVID-19 hotspot counties—United States, April–September 2020. MMWR Morb. Mortal. Wkly Rep. 69.

Pucher, J., Buehler, R., 2016. Safer Cycling Through Improved Infrastructure. American Public Health Association.

Pucher, J., Renne, J.L., 2003. Socioeconomics of urban travel. Evidence From the 2001 NHTS.

R Core Team, 2018. R: a language and environment for statistical computing. https://www.r-project.org/.

Rojas-Bolivar, D., Intimayta-Escalante, C., Cardenas-Jara, A., Jandarov, R., Huaman, M.A., 2021. COVID-19 case fatality rate and tuberculosis in a metropolitan setting. J. Med. Virol. 93 (5), 3273–3276.

Sanna, M., Hsieh, Y.-H., 2017. Ascertaining the impact of public rapid transit system on spread of dengue in urban settings. Sci. Total Environ. 598, 1151–1159.

Scarpone, C., Brinkmann, S.T., Große, T., Sonnenwald, D., Fuchs, M., Walker, B.B., 2020. A multimethod approach for county-scale geospatial analysis of emerging infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany. Int. J. Health Geogr. 19 (1), 1–17.

Schneider, R.J., Vargo, J., Sanatizadeh, A., 2017. Comparison of US metropolitan region pedestrian and bicyclist fatality rates. Accid. Anal. Prev. 106, 82–98.

Sharifi, A., Khavarian-Garmsir, A.R., 2020. The COVID-19 pandemic: impacts on cities and major lessons for urban planning, design, and management. Sci. Total Environ., p. 142391

Shen, J., Duan, H., Zhang, B., Wang, J., Ji, J.S., Wang, J., Pan, L., Wang, X., Zhao, K., Ying, B., et al., 2020. Prevention and control of COVID-19 in public transportation: experience from China. Environ. Pollut. 115291.

Stephanie, B., Virginia, B., Nancy, C., Aaron, C., Ryan, G., Aron, H., Michelle, H., Tamara, P., Matthew, R., 2020. Geographic differences in COVID-19 cases, deaths, and incidence-United States, February 12-April 7, 2020. MMWR Morb. Mortal. Wkly Rep. 69.

Sy, K.T.L., Martinez, M.E., Rader, B., White, L.F., 2020. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. Am. J. Epidemiol. 190 (7), 1234–1242.

Szklo, M., Nieto, F.J., 2014. Communicating results of epidemiologic studies. Epidemiology: Beyond the Basics. Jones & Bartlett, pp. 384–386.

Teixeira, J.F., Lopes, M., 2020. The link between bike sharing and subway use during the COVID-19 pandemic: the case-study of New York's citi bike. Transp. Res. Interdiscip. Perspect. 6, 100166.

Tirachini, A., Cats, O., 2020. COVID-19 and public transportation: current assessment, prospects, and research needs. J. Public Transp. 22 (1), 1.

U.S. Census Bureau, 2019. American community survey 5-year data API (2009-2019). https://www.census.gov/data/developers/data-sets/acs-5year.html.

U.S. Census Bureau, 2020. About Metropolitan and Micropolitan statistical areas. https://www.census.gov/programs-surveys/metro-micro/about.html.

U.S. Dept of Transportation, F.H.A., 2017. 2017 national household travel survey. https://nhts.ornl.gov.

Xu, F., McCluskey, C.C., Cressman, R., 2013. Spatial spread of an epidemic through public transportation systems with a hub. Math. Biosci. 246 (1), 164–175.

Yang, X., Duan, Q., Wang, J., Zhang, Z., Jiang, G., 2014. Seasonal variation of newly notified pulmonary tuberculosis cases from 2004 to 2013 in Wuhan, China. PLoS One 9 (10), e108369.

Yang, B., Wu, P., Lau, E.H.Y., Wong, J.Y., Ho, F., Gao, H., Xiao, J., Adam, D.C., Ng, T.W.Y., Quan, J., et al., 2021. Changing disparities in COVID-19 burden in the ethnically homogeneous population of Hong Kong through pandemic waves: an observational study. Clin. Infect. Dis. https://doi.org/10.1093/cid/ciab002.

Yashima, K., Sasaki, A., 2014. Epidemic process over the commute network in a metropolitan area. PLoS One 9 (6), e98518.

Zamudio, C., Krapp, F., Choi, H.W., Shah, L., Ciampi, A., Gotuzzo, E., Heymann, J., Seas, C., Brewer, T.F., 2015. Public transportation and tuberculosis transmission in a high incidence setting. PLoS One 10 (2), e0115230.

Zheng, R., Xu, Y., Wang, W., Ning, G., Bi, Y., 2020. Spatial transmission of COVID-19 via public and private transportation in China. Travel Med. Infect. Dis. 34, 1–3.