Super-Spreader Businesses and Risk of COVID-19 Transmission

Ashley O'Donoghue[1], Tenzin Dechen[1], Whitney Pavlova[2], Michael Boals[3], Garba Moussa[3], Manvi Madan[3], Aalok Thakkar[4], Frank J. DeFalco[5]

[1]*Center for Healthcare Delivery Science, Beth Israel Deaconess Medical Center, Boston, MA*
[2]*Department of Statistics, Pennsylvania State University, State College, PA*
[3]*Independent Researcher*
[4]*University of Pennsylvania, Philadelphia, PA*
[5]*Janssen Research & Development, Titusville, NJ*

Abstract

Purpose: The United States has the highest number of confirmed COVID-19 cases in the world to date, with over 94,000 COVID-19-related deaths[1]. The true risk of a COVID-19 resurgence as states prepare to reopen businesses is unknown. This paper aims to classify businesses by their risk of transmission and quantify the relationship between the density of super-spreader businesses and COVID-19 cases.

Methods: We constructed a COVID-19 Business Transmission Risk Index based upon the frequency and duration of visits and square footage of businesses pre-pandemic in 2019 in 8 states (Massachusetts, Rhode Island, Connecticut, New Hampshire, Vermont, Maine, New York, and California). We used this index to classify businesses as super-spreaders. Then, we analyzed the association between the density of super-spreader businesses in a county and the rate of COVID-19 cases. We performed significance testing using a negative binomial regression. The main outcome of interest is the cumulative number of COVID-19 cases each week.

Results: We found a positive association between the density of super-spreader businesses and COVID-19 cases. A 1 percentage point increase in the density of super-spreader businesses is associated with 5% higher COVID-19 cases, all else equal.

Correspondence: Ashley O'Donoghue, Center for Healthcare Delivery Science, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215.
E-mail: aodonogh@bidmc.harvard.edu

Conclusion: Higher densities of super-spreader businesses are associated with higher rates of COVID-19 cases. This may have important implications for how states reopen super-spreader businesses.

Keywords: COVID-19, Public Health, Epidemiology, Businesses, Super-spreader, Transmission

1. Introduction

The United States has the highest number of confirmed COVID-19 cases in the world to date, with over 94,000 COVID-19-related deaths[1]. A primary reason has been the emergence of various clusters of COVID-19 transmission from super-spreader events and establishments[2][3][4][5][6][7]. Identifying potential super-spreader businesses has important implications for policy-makers as they decide when and how to safely reopen non-essential businesses[8][9]. Baicker et al aimed to determine which industries or business establishments had a higher risk of transmission. The study raised important questions that individuals may face as businesses reopen, including the comparative risk of visiting different business establishments[7].

There is a pattern to the events and places that have a high risk of transmission that can be deemed as super-spreaders. They are often indoor events with people in extremely close proximity to each other for a long duration of time. The risk of transmission in a closed establishment is 18.7 times higher than in an event in an open-air establishment[10]. Even though the public and states are ready to to reopen the economy, experts cautioned on resurgence of the virus and death tolls if we open our economy prematurely[11][12][13][14][15].

Given the empirical evidence on the potential impact of super-spreaders in the spread of COVID-19, it is crucial to evaluate which businesses, events, establishments, and industries should reopen first and which ones may have a higher risk of spreading the virus. In this study, we sought to identify the businesses that have the potential to be super-spreaders before reopening the business establishments in different counties to aid the decision-making process for the policymakers. We tested the hypothesis that US counties with higher densities of super-spreader businesses, as defined by our index, were at a higher risk of COVID-19 transmission and thus may require a careful reopening of businesses to minimize a resurgence of COVID-19 cases.

2. Data and Methods

2.1 Data

We use data from SafeGraph Monthly Patterns[0] from January 1, 2019 – May 31, 2019 and SafeGraph Core Points of Interest data to measure business characteristics and

[0]SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block

traffic. Data on county-level COVID-19 cases and tests are from Johns Hopkins University and the New York Times. Socio-economic and demographic characteristics are collected from the 2018 American Community Survey from the United States Census Bureau. Businesses are classified by their 6-digit North American Industry Classification System (NAICS) code, developed by the United States Census Bureau.

2.2 Setting

This study focused on counties in 8 states (Massachusetts, Rhode Island, Connecticut, New Hampshire, Vermont, Maine, New York, and California). There are 187 counties, with a total population of 73,894,989. We examine traffic to 918,094 businesses from 307 different 6-digit NAICS codes from January 1, 2019 – May 31, 2019. We analyze COVID-19 cases in these counties from January 22, 2020 - May 22, 2020.

2.3 Index Construction and Super-Spreader Classification

We constructed a COVID-19 Business Transmission Risk Index using data on business characteristics and traffic by NAICS code from January 1, 2019 – May 31, 2019. The index was built using data on visitors per square foot, frequency of visits, and the average duration of visits. Visitors per square foot account for how densely visitors are packed into businesses. Businesses that are more densely packed may have a higher risk of COVID-19 transmission. The average duration of visits accounts for the length of time visitors are spending in a business. Businesses where visitors linger for longer periods of time could be riskier for COVID-19 transmission than businesses where visitors are quickly in and out of the business. The COVID-19 Business Transmission Risk Index is calculated for each 6-digit NAICS code in our sample by weighting the total visit time across all visitors from January 1, 2019 – May 31, 2019 by the square footage of the business establishment. NAICS codes which fall in the top 5% of the Index are classified as super-spreader industries. We classify businesses in these industries as super-spreader businesses. This classifies 156,307 individual businesses as super-spreaders out of a total of 918,094 businesses.

2.4 Study Variables

The outcome measure is the cumulative number of COVID-19 cases each week per county. The independent variable is the density of super-spreader businesses in a county, which is measured as the number of super-spreader businesses out of the total number of businesses. Covariates included are counties' racial composition (Black and White), population above 65 years, population below the poverty line, and population density per square mile.

group.

Table 1. Summary Statistics

COVID-19 Cases per 10,000	28.83 (53.36)
Super-spreader Businesses Density	16.29 (3.74)
% Over Age 65	18.55 (4.61)
% White	84.86 (11.30)
% Black	2.95 (3.45)
% Below Federal Poverty Line	8.96 (3.41)
Population Density	299.49 (656.53)
Total Population	73,894,989
Total Businesses	918,094
Number of Counties	187
Number of States	8

Means with standard deviations in parentheses.

2.5 Statistical Analysis

Univariate analyses were conducted to produce overall baseline characteristics and most common super-spreader businesses. Data were analyzed using a negative binomial regression at the county-level. The natural log of the total county population was included as an offset term. The model was adjusted for counties' racial composition, percent of population above 65 years, percent of population below the poverty line, and population density. Additionally, an indicator variable for each state was included in an effort to adjust for differences in testing practices across states, though this will not account for differences in testing across states that varies over time. Standard errors were clustered at the state level.

Coefficients were transformed into incidence rate ratios (IRRs) and are reported with 95% confidence intervals (CIs). Statistical significance was determined by a p-value ≤ 0.05. All tests were two-tailed. Statistical analysis was performed using Stata SE version 14.2 (StataCorp).

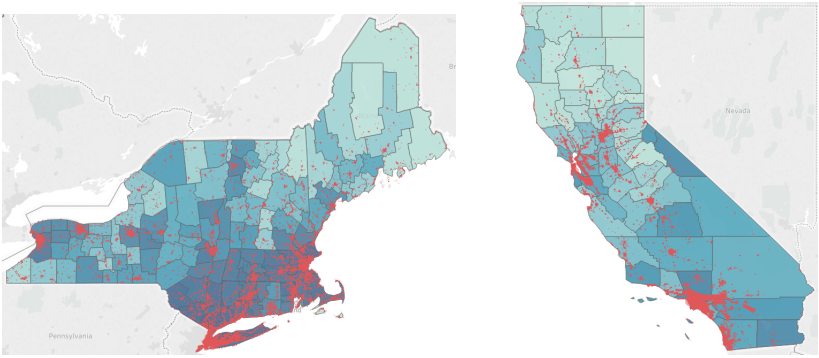
3. Results

3.1 Summary Statistics

Summary statistics are reported in Table 1. In our sample, there were an average 28.83 cumulative cases of COVID-19 per 10,000 by May 22, 2020. The average density of super-spreader businesses in a county was 16.29 per 100 businesses. On average, 18.55% of a county was above the age of 65, 84.86% was White, 2.95% was Black, and 8.96% was below the federal poverty line. The average population density of a county was 299.49 people per square mile. Our study covered 8 states, 187 counties, with a total population of 73,894,989, and 918,094 businesses.

Figure 1 displays a map of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York and California, with total cumulative COVID-19

cases for each county as of May 22, 2020 and locations of super-spreader businesses.



(a) Super-spreader businesses and COVID-19 cases per 10,000 in the Northeast.

(b) Super-spreader businesses and COVID-19 cases per 10,000 in California.

Fig. 1. The color density of the plots are based on a percentile rank of the total COVID-19 case rates for all counties in the study, ranging from 0 to 1,203.6 cases per 10,000 people. Super-spreader businesses are also displayed on the map as red dots with their relative size reflecting the estimated total dwell time for people in each location.

3.2 Main Results

Table 2 reports the most common super-spreader business types by NAICS code. The most common type of super-spreader business in our sample is full-service restaurants. These are restaurants where you're seated, typically have a server, and pay after your meal is completed. There are 116,605 full-service restaurants in our sample. The second most common type of super-spreader business is limited-service restaurants with 26,196 in our sample. These are restaurants where you may pay at a counter prior to your meal. This would include fast food, delicatessens, sandwich shops, take-out restaurants, and pizza delivery. The third most common type of super-spreader business in our sample is hotels (except casino hotels) and motels with 13,432 of these businesses in our sample.

Table 3 reports the main results of our negative binomial regression measuring the association between super-spreader density and COVID-19 cases. In table 3, $p <$

Table 2. Most Common Super-spreader Business Types

Business Types	Count of Businesses in Sample
Full-Service Restaurants	116,605
Limited-Service Restaurants (ex: fast food, delis, takeout)	26,196
Hotels (except Casino Hotels) and Motels	13,432

0.01. The estimates are incidence rate ratios, weighted by total county population, adjusted for population over age 65, racial distribution, population below the poverty line, and population density. Includes an indicator variable for each state. Standard errors clustered at the state-level. We find a positive association between the density of super-spreader businesses and COVID-19 cases (adjusted IRR=1.05; 95% CI: $1.02 - 1.07$). Our results suggest that an increase in super-spreader businesses by 1 percentage point results in a 5% increase in COVID-19 cases, all else equal.

Table 3. Main Results

	COVID-19 Cases IRR/(95% CI)
Super-spreader Business Density	1.05*** (1.02 – 1.07)
Number of Observations	1,060

*** $p < 0.01$.

Estimates are incidence rate ratios. Weighted by total county population.

Adjusted for population over age 65, racial distribution, population below the poverty line, and population density. Includes an indicator variable for each state.

Standard errors clustered at the state-level.

4. Discussion

4.1 Super-spreader Businesses and COVID-19

Our index attempts to quantify the risk of COVID-19 transmission at businesses based upon the frequency and duration of visits as well as the density of visitors in the businesses. Businesses with more visitors that stay for longer and are more densely packed are likely to have higher risks of transmission.

Knowing the density of super-spreader businesses will be very useful for policymakers. This can allow policy-makers to help plan to reopen these super-spreader businesses in the safest way possible. Our index classifies restaurants as the most common type of super-spreader business. When planning to reopen, policymakers can consider more options to help restaurants reopen while mitigating the risk to the public. This could include more outside seating, limitations on the number of visitors at a time, and monitoring traffic to potential super-spreader businesses.

This study can also be useful for hospital decision-makers. Knowledge of the density of super-spreader businesses and monitoring traffic to these businesses may help hospitals prepare for a potential second-wave if traffic increases to these businesses very quickly.

4.2 Limitations

There are several limitations to this study. First, COVID-19 cases are based upon a positive COVID-19 test. Thus, this will not account for individuals who may be COVID-19 positive but did not receive a test, either because of scarcity of tests or because they were asymptomatic. To help mitigate some of this bias, we also plan to explore other measures of COVID-19 incidence at the county-level, such as the percent of total tests that are positive.

Second, while we control for population density at the county-level, there is variation in population density within counties that is likely correlated with both the variation in super-spreader business density and COVID-19 cases within counties. Thus, we are currently seeking out more granular data on COVID-19 cases in order to more accurately adjust for potential confounding by population density.

4.3 Future Work

We are continuing to work on this study. First, we plan to incorporate all 50 states into the next iteration of our analysis. Second, we also plan to incorporate business airflow into the Index, such as outside seating options for restaurants.

Third, we are currently building an online decision-support tool that will allow policy-makers and hospital decision-makers to visualize potential super-spreader businesses in their area and monitor weekly traffic to these businesses. This can help policy-makers and hospital decision-makers plan for a potential second wave.

Finally, as states begin to reopen non-essential in phases, we plan to evaluate the effects of these reopenings on COVID-19 transmission. We plan to implement a difference-in-differences event study framework to measure the dynamic effects of reopening on COVID-19 cases. Knowing the effects of reopening can help future policy-makers and hospital decision-makers plan for the potential impact of reopening.

5. Conclusion

In conclusion, we built a COVID-19 Business Transmission Risk Index based on the frequency, density, and duration of visitors to businesses in 8 states. We find a positive association between the density of super-spreader businesses and COVID-19 cases in a county. We control for several socio-demographic and economic characteristics of counties, population density, and attempt to account for differences in testing across states.

This study can have important implications for policymakers as they consider how to most safely reopen these potential super-spreader businesses. We continue to work on acquiring more granular data to better account for confounding from population density. We also are in the process of building a tool for policymakers and

hospital decision-makers to monitor traffic to potential super-spreader businesses in their community as they begin to reopen.

References

1. Coronavirus disease (COVID-19) Situation Report – 125. Tech. rep. World Health Organization, 2020; Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200524-covid-19-sitrep-125.pdf.
2. Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for SARS-CoV-2. Lancet, Mar. 2020;395(10227): e47.
3. Woodward A. Coronavirus super-spreader events all have notable similarities — and they reveal the types of gatherings we should avoid for years. Business Insider, 2020; Available from: https://www.businessinsider.com/coronavirus-super-spreader-events-reveal-gatherings-to-avoid-2020-5.
4. Williamson E, Hussey K. Party Zero: How a Soirée in Connecticut Became a 'Super Spreader'. The New York Times, 2020; Available from: https://www.nytimes.com/2020/03/23/us/coronavirus-westport-connecticut-party-zero.html.
5. Stockman F, Barker K. How a Premier U.S. Drug Company Became a Virus 'Super Spreader'. The New York Times, 2020; Available from: https://www.nytimes.com/2020/04/12/us/coronavirus-biogen-boston-superspreader.html.
6. Correa-Mart?nez CL, Kampmeier S, K?mpers P, Schwierzeck V, Hennies M, Hafezi W, et al. A pandemic in times of global tourism: superspreading and exportation of COVID-19 cases from a ski area in Austria. J. Clin. Microbiol. 2020;
7. Al-Tawfiq JA, Rodriguez-Morales AJ. Super-spreading events and contribution to transmission of MERS, SARS, and COVID-19. J. Hosp. Infect. 2020;
8. Baicker K, Dube O, Mullainathan S, Pope D, Wezerek G. Is It Safer to Visit a Coffee Shop or a Gym? The New York Times, Available from: https://www.nytimes.com/interactive/2020/05/06/opinion/coronavirus-us-reopen.html.
9. Benzell S, Collis A, Nicolaides C. Rationing Social Contact During the COVID-19 Pandemic: Transmission Risk and Social Benefits of US Locations. The New York Times, Available from: httpshttps://ssrn.com/abstract=3579678.
10. Hiroshi N, Hitoshi O, Tetsuro K, Tomoya S, Tomimasa S, Tamano M, et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). medRxiv, 2020; eprint: https://www.medrxiv.org/content/early/2020/04/16/2020.02.28.20029272.full.pdf. Available from: https://www.medrxiv.org/content/early/2020/04/16/2020.02.28.20029272. doi: 10.1101/2020.02.28.20029272.
11. Samuel J, Rahman MM, Ali GGMN, Samuel Y, Pelaez A. Feeling Like It Is Time to Reopen Now? COVID-19 New Normal Scenarios Based on Reopening Sentiment Analytics. Preprint, 2020; 2020050318. Available from: doi:10.20944/preprints202005.0318.v1.
12. Dhillon RS, Karan A, Beier D, Sullivan A, Chowell G, Chowell D, et al. A Plan to Safely Reopen the U.S. Despite Inadequate Testing. Tech. rep. Harvard Business Review, 2020; Available from: https://hbr.org/2020/05/a-plan-to-safely-reopen-the-u-s-despite-inadequate-testing.
13. Killian JA, Charpignon M, Wilder B, Perrault A, Tambe M, Majumder MS. Evaluating COVID-19 Lockdown and Reopening Scenarios For Georgia, Florida, and Mississippi. SSRN, 2020; Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3598744.
14. Yamana T, Pei S, Kandula S, Shaman J. Projection of COVID-19 Cases and Deaths in the US as Individual States Re-open May 4,2020. medRxiv, 2020; eprint: https://www.medrxiv.org/content/early/2020/05/13/2020.05.04.20090670.full.pdf. Available from: https://www.medrxiv.org/content/early/2020/05/13/2020.05.04.20090670. doi: 10.1101/2020.05.04.20090670.
15. Angulo FJ, Finelli L, Swerdlow DL. Reopening Society and the Need for Real-Time Assessment of COVID-19 at the Community Level. JAMA, May 2020; ISSN: 0098-7484. eprint: https://jamanetwork.com/

journals/jama/articlepdf/2766293/jama_angulo_2020_vp_200096.pdf. Available from: https://doi.org/10.1001/jama.2020.7872. doi: 10.1001/jama.2020.7872.