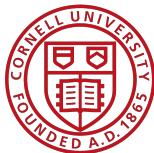# The Impact of Crowded Housing on COVID-19 Transmission Dynamics in New York City

Sara Venkatraman

Cornell University, Statistics and Data Science

Introduction

Hypothesis: People in overcrowded and/or multigenerational households may be at increased risk of COVID-19 infection

nature medicine

Letter | Published: 05 October 2020

Crowding and the shape of COVID-19 epidemics

Benjamin Rader, Samuel V. Scarpino ✉, Anjalika Nande, Alison L. Hill, Ben Adlam, Robert C. Reiner, David M. Pigott, Bernardo Gutierrez, Alexander E. Zarebski, Munik Shrestha, John S. Brownstein, Marcia C. Castro, Christopher Dye, Huaiyu Tian, Oliver G. Pybus ✉ & Moritz U. G. Kraemer ✉

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2

Kaiyuan Sun*†, Wei Wang†, Lidong Gao†, Yan Wang, Kaiwei Luo, Lingshuang Ren, Zhifei Zhan, Xinghui Chen, Shanlu Zhao, Yiwei Huang, Qianlai Sun, Ziyan Liu, Maria Litvinova, Alessandro Vespignani, Marco Ajelli, Cécile Viboud‡, Hongjie Yu*‡

The New York Times

'We Are Forced to Live in These Conditions': In Los Angeles, Virus Ravages Overcrowded Homes

Our work: ZIP code-level analysis of association between household composition and COVID-19 case counts in NYC

Our data

COVID-19 (suspected) cases: From NYC Dept. of Health.

- Daily count (over March 2020) of suspected COVID-19 cases in each NYC ZIP code, presenting to one of 53 emergency depts.

**Household composition data:** From Census Bureau.

- % of each ZIP's homes that are overcrowded or multigenerational

- "Overcrowded": > 1 occupant per room

- "Multigenerational": grandparents living with grandchildren

Socioeconomic and clinical factors: From Census Bureau and CDC. For each ZIP:
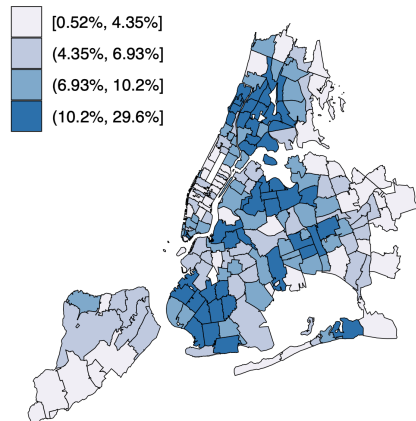
- Median income, % of essential workers, racial composition

- Prevalence of obesity, hypertension, diabetes, smoking

Spatial distribution of suspected COVID-19 cases
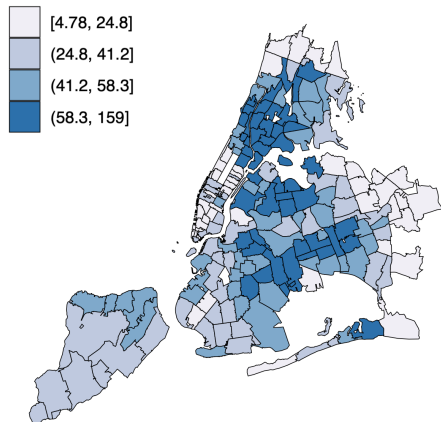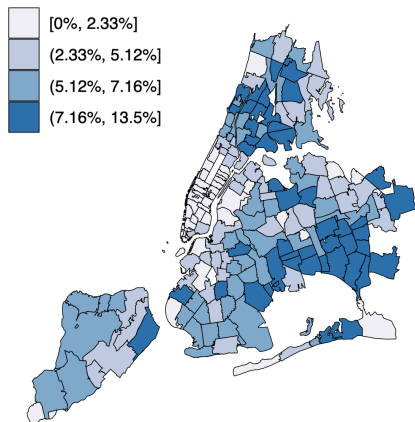

Suspected COVID−19 Cases Per 10,000 People

- [4.78, 24.8]
- (24.8, 41.2]
- (41.2, 58.3]
- (58.3, 159]
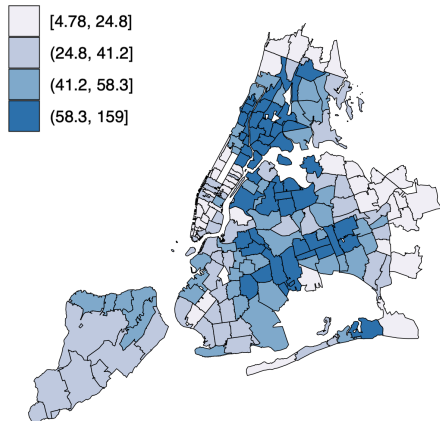
Percentage of Overcrowded Homes

- [0.52%, 4.35%]
- (4.35%, 6.93%]
- (6.93%, 10.2%]
- (10.2%, 29.6%]

Spatial distribution of suspected COVID-19 cases

Suspected COVID−19 Cases Per 10,000 People

- [4.78, 24.8]
- (24.8, 41.2]
- (41.2, 58.3]
- (58.3, 159]

Percentage of Multigenerational Homes

- [0%, 2.33%]
- (2.33%, 5.12%]
- (5.12%, 7.16%]
- (7.16%, 13.5%]

Modeling with Poisson regression

Poisson regression model of suspected COVID-19 case counts in each ZIP over time as a function of several covariates:

$$\log \begin{pmatrix} \text{Case count in} \\ \text{ZIP } i \text{ at time } t \end{pmatrix} = \beta_0 + \beta_1 t + \beta_2 C_i + \beta_3 M_i + \begin{pmatrix} \text{Other variables} \\ \text{for ZIP } i \end{pmatrix}$$

- t = time in days since March 1, 2020
- C_i = % of homes that are overcrowded in ZIP i
- M_i = % of homes that are multigenerational in ZIP i
- Other variables for ZIP i:
  - % of ZIP i with heart disease, obesity, smoking habits
  - ZIP i's median income, % essential workers, racial/ethnic composition

Modeling with Poisson regression (continued)

However, standard Poisson regression assumes case counts are independent of each other in space and time.

We can account for spatio-temporal structure by modeling ZIP case count as:

- Conditionally autoregressive in space (a "Besag-York-Mollié" model)
- A random walk in time

Poisson regression models with these assumptions can be fit in R with the INLA package.

Poisson spatio-temporal modeling

General form of Besag-York-Mollié model, with a temporal component:

$$\log \left(\begin{smallmatrix} \text{Case rate in} \\ \text{area } i \text{ at time } t \end{smallmatrix} \right) = \beta_0 + \mathbf{X}_i \boldsymbol{\beta} + u_i + \gamma_t + v_i + \phi_t$$

- $\beta_0 =$ average case rate across all areas
- $\mathbf{X}_i \boldsymbol{\beta} =$ vector of explanatory spatial covariates (e.g., ZIP code's % of overcrowded/multigenerational housing, median income, etc.)
- $u_i =$ conditionally autoregressive spatial effect. Normally distributed with mean $\frac{1}{|N_i|} \sum_{j \in N_i} u_j$, where $N_i =$ indices of neighboring areas
- $\gamma_t =$ temporal effect modeled as random walk: $\gamma_t - \gamma_{t-1}$ is normally distributed with mean 0
- $v_i, \phi_t =$ unstructured residuals accounting for other variation in space/time
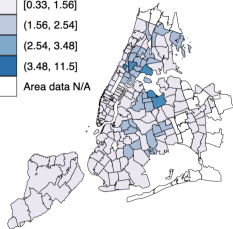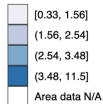
Model results

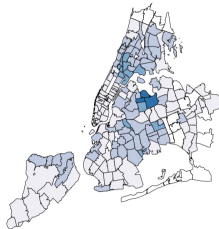**Trend in COVID–19 cases by quartile of overcrowding**
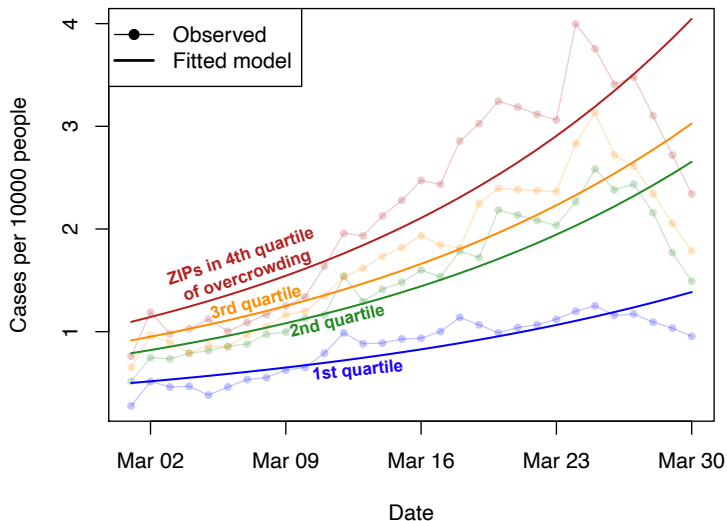
Model results (continued)
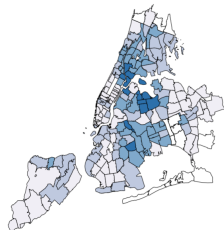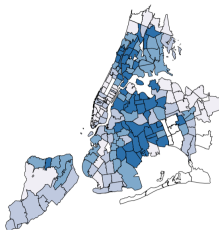
Conclusions

- Increased proportions of households that are overcrowded and/or multigenerational are associated with increased infection rates, after accounting for relevant risk factors

- Corroborates previous studies analyzing:
  - Racial and economic disparities in COVID-19 case rates
  - Correlation between infection and transmission proximity

- Ongoing work: examining the impact of school closures and social distancing orders on infection rates in crowded neighborhoods

Thank you!

Sara Venkatraman

skv24@cornell.edu
https://sara-venkatraman.github.io