Understanding the Los Angeles County Coronavirus Epidemic: The Critical Role of Intrahousehold Transmission

Jeffrey E. Harris MD PhD*† October 12, 2020

Abstract: We tracked the course of the COVID-19 epidemic among the approximately 300 communities comprising Los Angeles County. The epidemic, we found, had three distinct phases. During Phase I, from early March through about April 4, initial seeding of infection in relatively affluent areas was followed by radial geographic extension to adjoining communities. During Phase II, which continued to about July 11, COVID-19 cases continued to rise at a slower rate, and became increasingly concentrated in four geographic foci of infection across the county. This phase saw changes in two indicators of social mobility: the proportion of location-tracked smartphones staying at home, and the number of smartphones visiting a gym. For both indicators, those communities with a larger reduction in social mobility during April reported fewer new COVID-19 cases in May. During Phase III, COVID-19 incidence only gradually declined, remaining as high as the incidence seen at the end of Phase I. Across communities, the prevalence of households at high risk for intergenerational transmission was strongly correlated with the persistence of continued COVID-19 propagation. This association was even stronger in those communities with a higher rate of gym attendance in Phase II. The map of the prevalence of at-risk households in Los Angeles County coincided strikingly with the map of cumulative COVID-19 incidence. These findings, taken together, support the critical role of household structure in the persistent propagation of COVID-19 infections in Los Angeles County. Public health policy needs to be reoriented from a focus on protecting the individual to a focus on protecting the household.

* Professor of Economics, Emeritus, Massachusetts Institute of Technology, Cambridge MA 02139; and Physician, Eisner Health, Los Angeles CA 90015. *jeffrey@mit.edu*

† This study relies exclusively on publicly available data that contain no individual identifiers. The author has no competing interests and no funding sources to declare. This article represents the sole opinion of its author and does not necessarily represent the opinions of the Massachusetts Institute of Technology, Eisner Health, the National Bureau of Economic Research, or any other organization. We gratefully acknowledge the assistance of Douglas Morales MPH and Rashmi Shetgiri MD MSHS of the Los Angeles County Department of Public Health.

Introduction

In this article, we attempt to identify the critical forces driving the massive outbreak of COVID-19 in Los Angeles County, which has to date registered over 266,000 confirmed cases and almost 6,400 deaths (Los Angeles County Department of Public Health 2020a).

To that end, we bring together four critical strands of the growing research literature on the worldwide COVID-19 epidemic. First, investigators have attempted to reconstruct the transmission dynamics of local outbreaks by applying theoretical models to data on reported cases (Hao et al. 2020, Fang, Nie, and Penny 2020). Second, cross-sectional studies have related the age structure and household composition of various countries to COVID-19 incidence and mortality (Aparicio and Grossbard 2020, Esteve et al. 2020). Third, numerous studies have used the techniques of geospatial analysis to evaluate the impacts of public health policies (Franch-Pardo et al. 2020, Dickson et al. 2020, Orea and Alvarez 2020). And fourth, researchers have increasingly relied on data derived from the movements of devices with location-tracking software to study patterns of viral propagation (Dave et al. 2020, Harris 2020c, h).

Here, we rely upon detailed data on the dynamics of COVID-19 transmission among approximately 300 communities within Los Angeles County during March through September 2020. We use geospatial modeling and mapping techniques to study the radial spread of infection among these contiguous communities. We merge our geospatial data with information on two indicators of social mobility: the number of devices staying completely at home, and the number of devices visiting any one of nearly two thousand gyms and fitness centers in the county. We further overlay these data on a community-specific measure of the prevalence of households at high risk for intergenerational transmission.

Data and Methods

Los Angeles County Department of Public Health Data

We relied on reports of confirmed COVID-19 cases issued by the Los Angeles County Department of Public Health (DPH) and regularly updated on its web-based data dashboard (Los Angeles County Department of Public Health 2020b). At the global county level, the dashboard provided the full history of cumulative cases by date of testing from March 1 through September 19, 2020, the cutoff date for this study.

At the detailed geographic level, by contrast, the dashboard showed only the cumulative number of confirmed cases to date. Still, relying on an online archive of contemporaneous, daily press releases issued by the DPH (Los Angeles County Department of Public Health 2020c), we were able to reconstruct the history of cumulative cases at detailed geographic level from March 21 onward. Prior to March 16, DPH did not report any geographic breakdown of cumulative case counts. We further excluded reports issued during March 16–20 because a substantial proportion of reported cases were still under investigation or their geographic unit of origin was withheld because of small numbers.

DPH employs a geographical breakdown based upon *countywide statistical areas* (CSAs), a mixed classification of independent cities such as the City of Beverly Hills, neighborhoods within the city of Los Angeles such as Hollywood, and unincorporated places such as Hacienda Heights (City of Los Angeles 2020). The DPH kindly provided a crosswalk between CSAs and census tracts, based upon the locations of the centroids of each census tract, which permitted us to link the CSAs to other census-tract-based data sources.

SafeGraph Data

We relied upon two data sources provided by SafeGraph: the Patterns database (SafeGraph Inc. 2020a), and the Social Distancing database (SafeGraph Inc. 2020b).

The Patterns database provides information on the movements of smartphones equipped with location-tracking software to numerous points of interest throughout the United States. We previously relied upon this data source in studies of visitors to President Trump's rally at Tulsa's BOK Center on June 20, 2020 (Harris 2020k), restaurant attendance in San Antonio around the time of street protests during May 30 – June 11, 2020 (Harris 2020j), and attendance at bars in a study of COVID-19 propagation in Milwaukee and Dane Counties in Wisconsin (Harris 2020c).

Here, we focused on a particular class of points of interest to young adults. Specifically, we used the Patterns *location_name* variable to assemble an extensive list of 1,995 gyms in Los Angeles County, including entities offering exercise and resistance training, Zumba, Pilates, yoga, boxing and martial arts, but excluding those offering only massage. For each gym, we used the variable *visitor_home_cbgs* to identify the home census block groups of all visitors during February through September 2020, where a device's *home* is the location where it is regularly located overnight. Taking advantage of the census tract crosswalk provided by the Los Angeles County DPH, we counted the number of device owners visiting gyms over time according to

their home CSA. As a result of a revision in the frequency of reporting of Patterns data, we ended up with two time series: the numbers of gym visitors originating from each CSA for each week from February 10 through April 27, 2020; and the numbers of gym visitors originating from each CSA for each month from May through August, 2020.

The Social Distancing database provides information on both the origin and destination census block groups of device holders. We previously relied upon this data source in a study of the movements of individuals with a *home* in the Queens-Elmhurst COVID-19 hot spot to subway stations in Queens and Manhattan during the earliest days of the epidemic in New York City (Harris 2020h).

Here, we focused on an indicator of social avoidance behavior more typical of older persons. Specifically, we used the variable *completely_home_device_count* to determine the numbers of devices originating in each census block group that stayed completely at home during a particular day. Using data on the corresponding *device_count* as a denominator for each census block group, and once again taking advantage of the census tract crosswalk provided by DPH, we constructed a time series of the mean percentage of devices staying completely at home in each CSA from the week starting February2 through the week starting September 13, 2020.

American Community Survey Data

We relied on the 2018 public use microsample from the U.S. Census Bureau's American Community Survey (ACS) (U.S Census Bureau 2018). The nationwide database covered 159,685 households and group living arrangements with a total of 378,817 persons. Out of the entire database, 41,086 households and group living arrangements with 102,092 persons were identified as residing in one of 69 public use microdata areas (PUMAs) within Los Angeles County (U.S Census Bureau 2020b).

We used the person records of the public use microsample to identify households with at least 4 persons, of whom at least one person was 18–34 years of age and at least one other person was at least 50 years of age. We describe these households here as *at risk* for intergenerational transmission. Among all such at-risk families in the Los Angeles County extract of the ACS, 44 percent had 4 persons, 27 percent had 5 persons, 14 percent had 6 persons, and 15 percent had 7 or more persons. Among the ACS's 69 public use microdata areas (PUMAs), the median proportion of at-risk households was 14.9 percent.

Using the internal household sampling weights provided by the ACS, we then computed the proportion of at-risk households in each PUMA. Applying a Census Bureau crosswalk between PUMAs and census tracts (U.S Census Bureau 2020a) and the DPH-provided crosswalk between census tracts and CSAs, we determined the corresponding proportions of at-risk households in each CSA. Among 300 CSAs, the median proportion of at-risk households was 14.6 percent. By way of sensitivity analysis, we also varied the definition of an at-risk household, changing the age- and number-of-person cutoffs.

Spatial Models

We devised a spatial adaptation of a parsimonious empirical SIR (susceptible-infective-resistant) model, similar to the model we employed in a study of COVID-19 transmission between younger and older persons in Florida's most populous counties (Harris 2020e).

To facilitate the exposition, we first review the continuous-time version of a parsimonious SIR model without spatial components (Harris 2020a). Let $S(t)$ denote the proportion of susceptible individuals at time t, $I(t)$ the proportion of infective individuals, and $R(t)$ the proportion of resistant individuals. We assume that $-\dot{S} = \alpha SI$, where $\dot{S} = dS/dt$ stands for the time derivative, and where α reflects the rate at which susceptible and infective individuals interact, as well as the likelihood that an interaction will result in transmission. We further assume that $\dot{I} = \alpha SI - \beta I$, where β is the rate at which infectives become resistant, either through recovery or death. Finally, the population is assumed closed, so that $\dot{S} + \dot{I} + \dot{R} = 1$. At the start of an epidemic where virtually the entire population is naïve to the infectious agent, we have $S \approx 1$, so that $-\dot{S}$, the rate of new infections, is effectively proportional to I, the stock of infectives.

The discrete-time, empirically implementable version of this parsimonious model is as follows. Let y_{it} denote the rate of new COVID-19 cases per unit population in geographic unit i during time interval t, where the geographic units refer to CSAs and the time intervals refer to successive weeks. Let X_{it} denote the corresponding stock of infective individuals per unit population in CSA i during week t. The equation of motion of X_{it} is given by $X_{it} = (1-\beta)X_{i,t-1} + y_{i,t-1}$, where β is the weekly depreciation rate of the stock of infectives. With a mean serial interval of 5.5 days (Griffin et al. 2020), we assume a weekly depreciation factor of $1 - \beta = exp(-7/5.5) = 0.28$, that is, $\beta = 72$ percent of current infectives become resistant each

week. We further assume that virtually all individuals are initially susceptible to infection. Given these assumptions, the model specification without spatial effects is $y_{it} = \alpha X_{i,t-1} + \varepsilon_{it}$, where the right-hand-side variable is constructed solely from prior weeks' incidence rates, where α is an unknown parameter, and where the error terms ε_{it} are independent and identically distributed.

We now consider the spatial adaptation of this parsimonious model. We specify

$$y_{it} = \alpha X_{i,t-1} + \gamma \sum_{j \neq i} w_{ij} X_{j,t-1} + \varepsilon_{it} \qquad (1),$$

where $\{X_{j,t-1}: j \neq i\}$ refers to the stocks of infectives in all other geographic units, where both α and γ are unknown parameters. Here, $\{w_{ij}\}$ are elements of a known symmetric matrix with zero diagonal elements, where each off-diagonal element represents the influence of geographic unit $j \neq i$ on the rate of new infections in unit i. To implement this model, we set the off-diagonal element $w_{ij} = 1$ if the distance between the centroids of CSA i and CSA j was 2 km or less, and $w_{ij} = 0$ otherwise, where the distances between centroids were calculated from the Haversine formula (Hedges 2002). We designate this specification as *Spatial Model 1*.

Strictly speaking, the theoretical SIR model restricts the constant term of the empirical Spatial Model 1 to be zero. That is, all new infections are assumed to arise from contact with other infective persons located within the same or adjacent geographic units. The principal limitation of this restriction is that the stock of infectives X_{it} may be measured with error. In particular, the reported counts of confirmed COVID-19 cases may significantly understate the actual numbers of infectives, especially when asymptomatic, infective persons do not seek testing. What's more, the strong assumption that the number undercounted infectives is proportional to the number of reported infectives may not be warranted. Accordingly, we consider a generalized version of our spatial model of the form

$$y_{it} = \mu + \alpha X_{i,t-1} + \gamma \sum_{j \neq i} w_{ij} X_{j,t-1} + \delta Z_{i,t-1} + \varepsilon_{it} \qquad (2),$$

where $Z_{i,t-1}$ represents a lagged exogenous characteristic of CSA i, and μ and δ are additional unknown parameters. We designate this specification as *Spatial Model 2*.

Results

Weekly Incidence of Confirmed COVID-19 Cases in Los Angeles County

Figure 1 plots the weekly incidence rate of confirmed COVID-19 cases per 100,000 population in Los Angeles County as a whole, derived from the DPH's data dashboard (Los Angeles County Department of Public Health 2020b). The data run from the week starting Sunday, March 1 to the week starting Sunday, September 20, 2020.

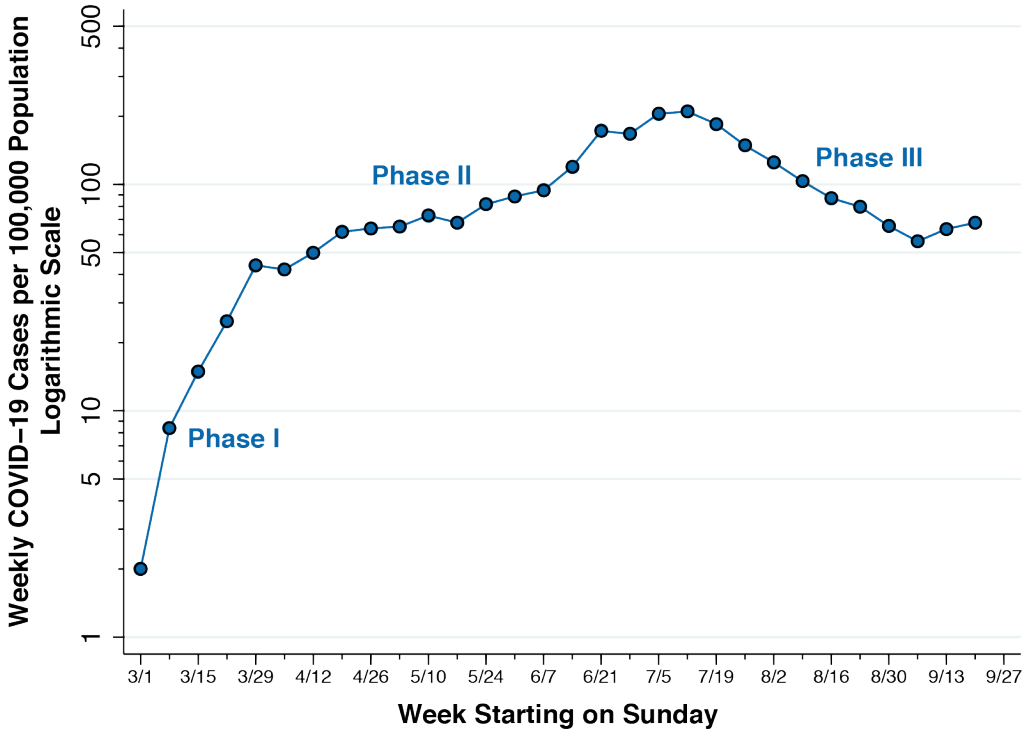


Figure 1. Weekly COVID-19 Cases per 100,000 Population in Los Angeles County

To facilitate our analysis, we have divided the observation period into three successive phases. *Phase I* represents an initial period of exponential growth through the week of March 29 – April 4. *Phase II* covers the subsequent period of continued slower growth, peaking around the week of July 5–11, while *Phase III* covers the final period of declining COVID-19 incidence.

Mapping the Geographic Spread of COVID-19 Across Countywide Statistical Areas

Appendix A shows 26 successive weekly maps of the cumulative incidence of confirmed COVID-19 cases per 100,000 during March 28 – September 19, 2020. The DPH did not report COVID-19 counts for the cities of Pasadena and Long Beach, which remain pale blue in each

map. The cumulative incidence across CSAs is color-coded in a geometric progression, starting with a threshold of 100 cases per 100,000 and successively doubling to 6,400 cases per 100,000.

Figure 2 below maps the status of the epidemic as of March 28 (that is, the end of the week starting March 22). By that point, the cumulative incidence had exceeded 100 cases per 100,000 in four relatively affluent CSAs: the Brentwood and Beverly Crest neighborhoods of Los Angeles, the City of West Hollywood, and the City of Palos Verdes Estates.

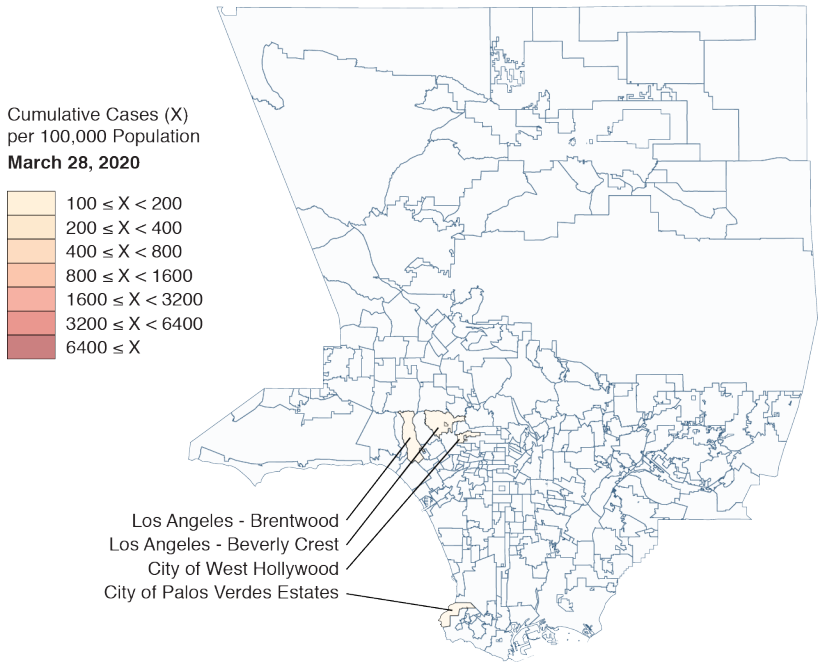


Figure 2. Cumulative Incidence of Confirmed COVID-19 Cases per 100,000 Population, Los Angeles County, as of March 28, 2020

The left and right panels of Figure 3 below, covering cumulative COVID-19 incidence through April 4 and 11, respectively, show the radial geographic expansion surrounding the initial focus of infection. By April 11, 2020, the cumulative incidence exceeded 200 per 100,000 in the adjacent neighborhoods of Beverly Crest, Melrose, Carthay, South Carthay, Crestview, Century City, as well as the neighboring City of Beverly Hills. The focus of infection had expanded to the northwest into the communities of Encino and Tarzana, to the west into the coastal community of Pacific Palisades and the City of Malibu, and to the north and east into Hollywood, East Hollywood, Silverlake, Glassell Park, and the City of Glendale.

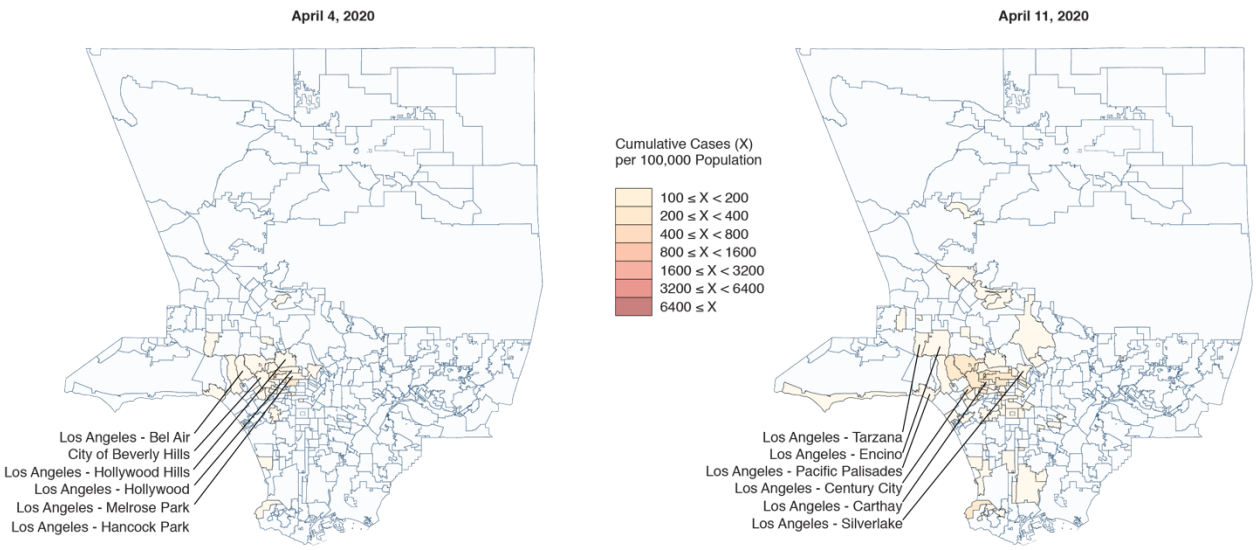


Figure 3. Cumulative Incidence of Confirmed COVID-19 Cases per 100,000 Population, Los Angeles County, as of April 4, 2020 (Left) and April 11, 2020 (Right)

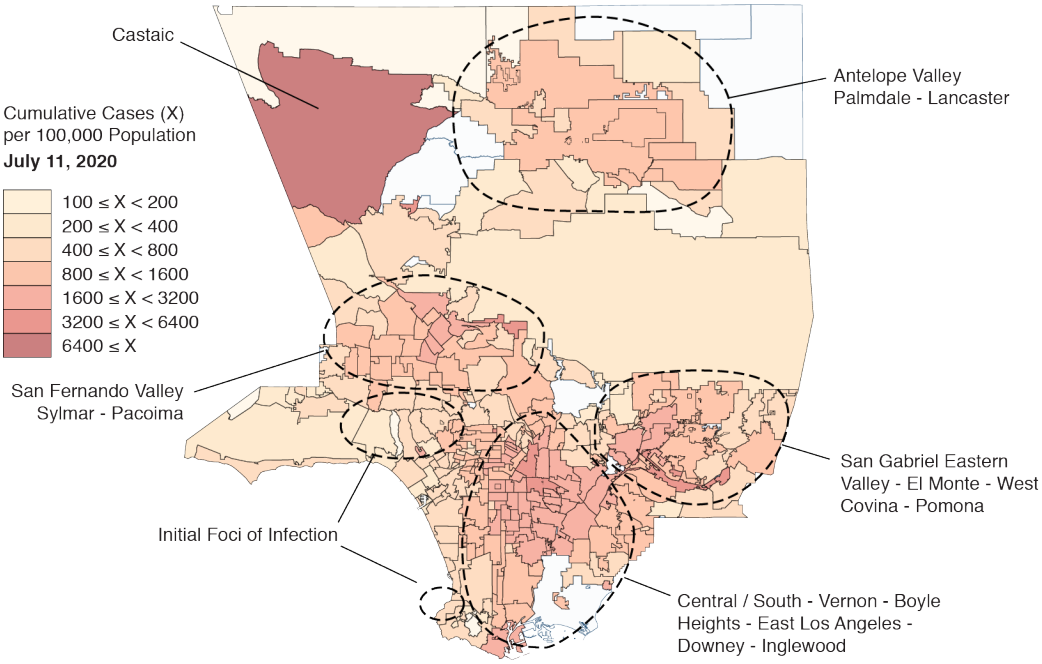


Figure 4. Cumulative Incidence of Confirmed COVID-19 Cases per 100,000 Population, Los Angeles County, as of July 11, 2020.

While the epidemic continued to expand radially during Phase II, as seen in Appendix A, four foci of infection began to emerge. Figure 4 above displays the epidemic map as of July 11 at the end of Phase II, when the overall weekly incidence of confirmed cases had reached its peak.

At the lower left, we see that relative few cases continued to accumulate at the two initial foci of infection seen in Phase I. At the upper left, we see a hot spot in the sparsely populated, unincorporated area of Castaic, which appears to have arisen from an outbreak in May among inmates and employees at the Pitchess Detention Center and its aftermath (Smith 2020, Tchekmedyian 2020). Aside from these identified areas, the map shows four new foci of infection: the Antelope Valley - Palmdale - Lancaster area to the north; the San Fernando Valley - Sylmar - Pacoima area to the west; the San Gabriel Eastern Valley - El Monte - West Covina - Pomona area to the east; and the Central/South - Vernon - Boyle Heights - East Los Angeles - Downey - Inglewood area in the center.

Figure 5 displays the epidemic map as September 19, the cutoff date of our study. The four principal foci of infection that emerged during Phase II have intensified, while the two initial foci identified in Phase I show relatively few cumulative cases per 100,000 population.

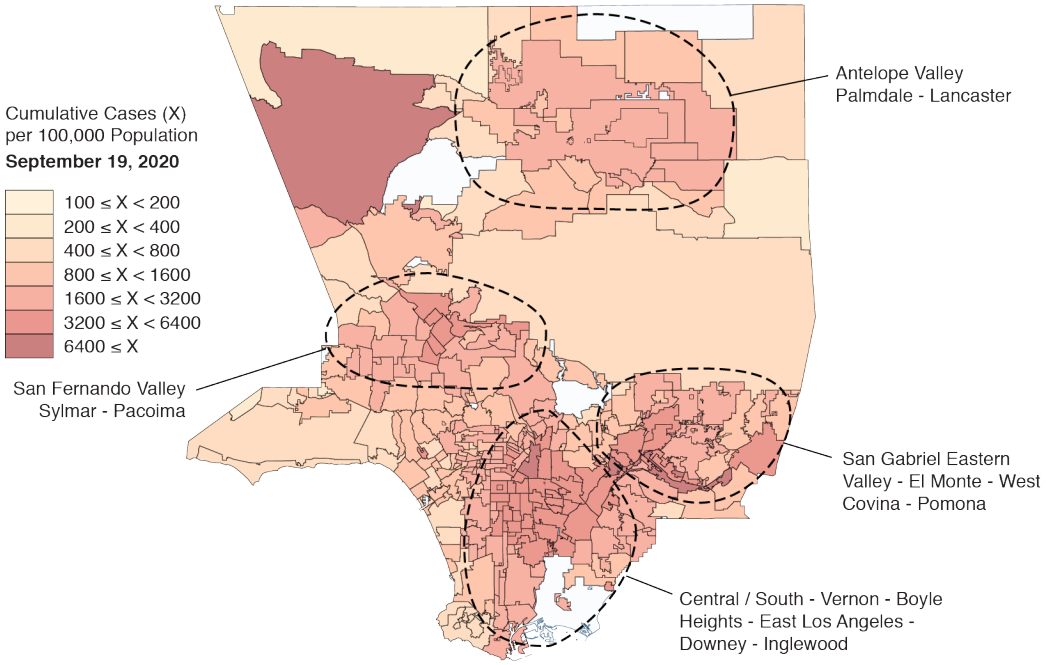


Figure 5. Cumulative Incidence of Confirmed COVID-19 Cases per 100,000 Population, Los Angeles County, as of September 11, 2020.

In summary, these geospatial observations support the conclusion that the Los Angeles County epidemic was initially seeded in two areas: Brentwood - Beverly Crest - West Hollywood to the west and Palos Verdes Estates to the south. During Phase I, which lasted until approximately April 4, new infections spread radially to adjoining geographic areas, and this

radial geographic extension continued at least during the first week of Phase II, ending on April 11. Thereafter, the epidemic appears to have been sustained not by continued radial expansion, but by the increasing concentration and growth of new infections in several identifiable foci. These foci of infection remained the dominant source of new COVID-19 cases even while the overall infection rate was declining in Phase III.

Tests of Spatial Model 1

Table B1 of Appendix B shows separate estimates of the parameters of Spatial Model 1 for the three phases of the epidemic. During Phase I, the effect of neighboring communities, as captured by the parameter γ (estimate 0.406, 95% confidence interval [CI] 0.299–0.513, $p < 0.001$), was 39 percent of the within-community effect, as captured by the parameter α (estimate 1.028, 95% CI 0.894–1.163, $p < 0.001$). This finding was consistent with the radial geographic spread observed in the mapping data. The estimated parameters imply a contemporaneous reproductive number of $(\alpha + \gamma)/\beta$ = 2.0 for the entire county, a value consistent with the deceleration of the incidence curve at the end of Phase I, as seen in Figure 1.

By contrast, during Phase II, the neighboring community effect γ (estimate 0.0.069, 95% confidence interval [CI] 0.054–0.083, $p < 0.001$), was only 12.4 percent of the within-community effect α (estimate 0.554, 95% CI 0.533–0.575, $p < 0.001$). This finding is consistent with map-based observations of reduced radial geographic propagation and continued within-CSA transmission during Phase II. The parameter estimates would imply a reproductive number of $(\alpha + \gamma)/(1 - \beta)$ = 0.86 for Los Angeles County as a whole. The continued growth in countywide case incidence seen in Figure 1, even in the face of a reproductive number below 1, reflects the concentration of new COVID-19 infections in the four principal foci identified in Figures 4 and 5.

During Phase III, the effect of neighboring communities, as captured by the parameter γ (estimate 0.027, 95% confidence interval [CI] -0.008–0.061, $p = 0.13$) was statistically indistinguishable from zero, while the within-community effect α (estimate 0.578, 95% CI 0.523–0.633, $p < 0.001$) was indistinguishable from the corresponding estimate for Phase II. This finding is consistent with the continued increasing concentration of COVID-19 infections within communities as overall incidence continued to decline.

Social Mobility Indicators: Percent of Devices Completely at Home

Figure 6 tracks the evolution of the percent of devices staying completely at home during the course of the epidemic in Los Angeles County. The timeline is marked in weeks, from the week starting Sunday February 2 to the week starting Sunday September 13. Four time series are shown, corresponding to CSAs in the four quartiles of the logarithm of cumulative COVID-19 incidence as of our cutoff date of September 19, 2020, where the first quartile has the highest cumulative incidence and the fourth quartile has the lowest incidence.

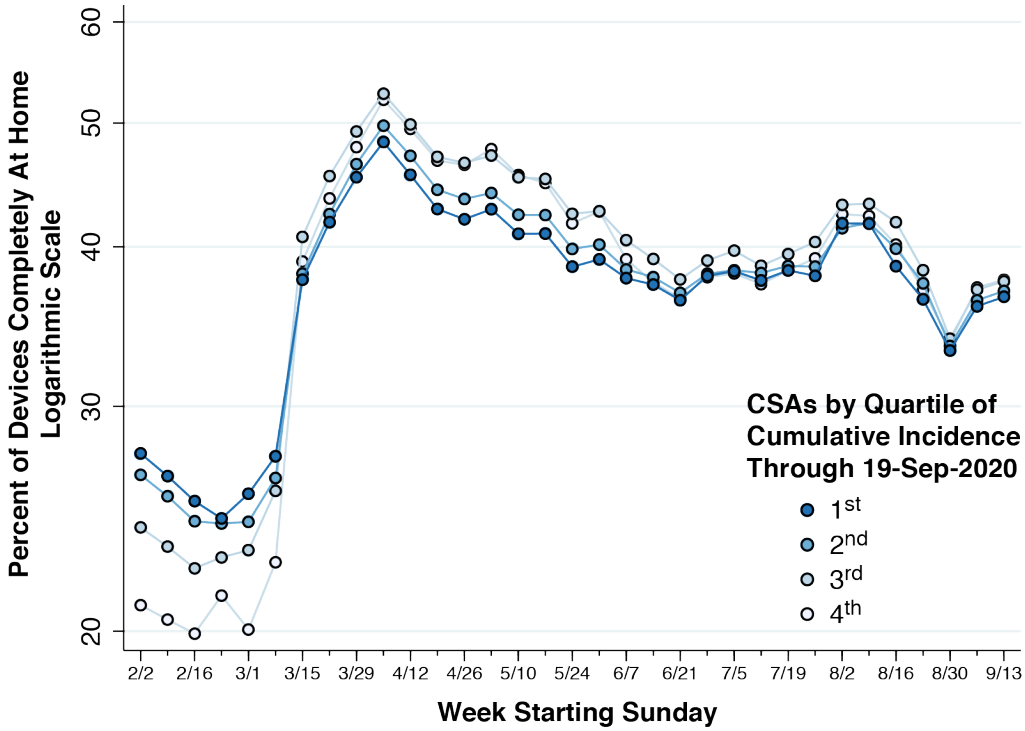


Figure 6. Percent of Devices Staying Completely at Home from Week of February 2 to Week of September 13, 2020, in Relation to the Quartile of Cumulative COVID-19 Incidence as of September 19, 2020

In the pre-epidemic period, during the weeks starting February 2 through February 23, there is clear gradient between the four quartiles. Those CSAs ultimately ending up with the most COVID-19 cases initially had the *largest* percentage of devices staying at home. Put differently, those geographic areas with the highest proportion of socially immobile persons at baseline ultimately suffered the greatest cumulative burden of viral infection. During epidemic Phases I through III, by contrast, the relationship between disease burden and social distancing reverses. Those CSAs that ultimately experienced the highest cumulative COVID-19 rates had

the *smallest* percent of devices staying at home. As discussed in more detail in Appendix C, this reversal was most pronounced during that portion of Phase I from March 15 onward, as well as in Phase II.

Table B2 in Appendix B shows the estimates of Spatial Model 2 with the inclusion of the percent of devices staying completely at home as a covariate. As anticipated from Figure 6, the regression results show a significant negative sign for the estimated parameter δ only during Phase II. That is, the higher the percentage staying completely at home, the lower was the incidence of newly confirmed COVID-19 infections, but only during Phase II.

Figure 7 below further explores the quantitative relationship between the percentage staying completely at home during the month of April and the subsequent number of newly confirmed COVID-19 infections during the subsequent month of May. Each data point is a single CSA. The legend at the left shows how the data points map into the four foci of infection identified in Figures 4 and 5. As indicated in Appendix D, the estimated population-weighted least squares regression line has a slope of –0.073, implying that an increase of 10 percentage points was associated with a 73-percent reduction in newly confirmed cases. These findings are likewise consistent with an impact of reduced social mobility during Phase II of the epidemic.

For reference, Appendix E shows the weekly incidence of newly confirmed COVID-19 infections separately for each focus of infection.

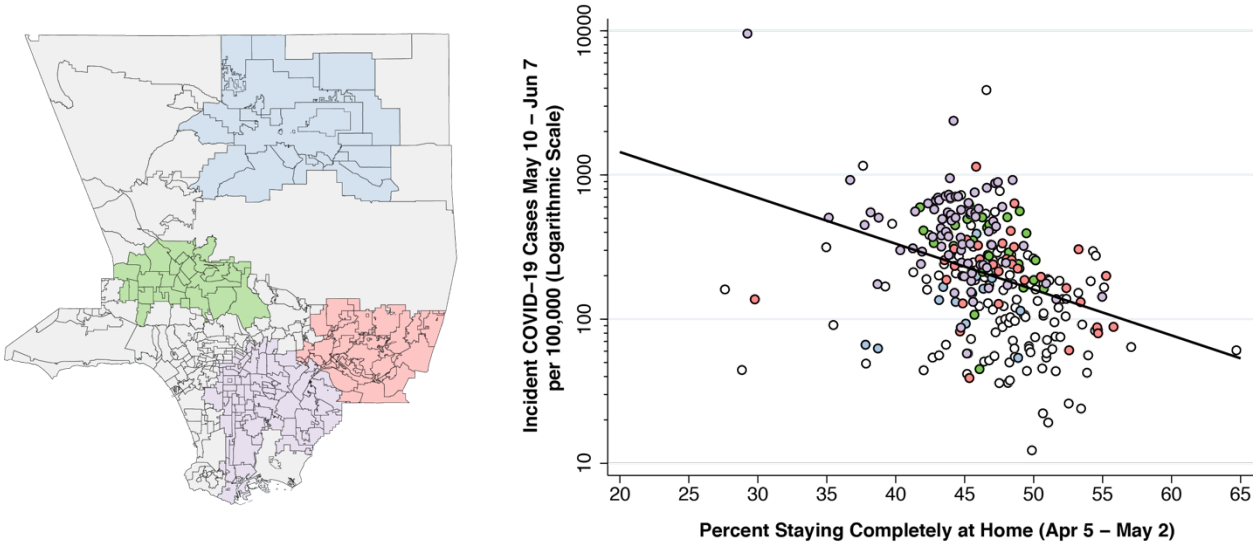


Figure 7. Relation Between Percent Staying Completely at Home During April 5 – May 2 and the Incidence of Newly Confirmed Cases During May 10 – June 7, 2020

Social Mobility Indicators: Visitors to Gyms

Figure 8 tracks the number of visitors to gyms from the week of Monday February 2 to the week of Monday April 27. The vertical axis has been normalized so that the week of March 2 equals 100. As in Figure 6, we show four time series corresponding to CSAs in the four quartiles of the logarithm of cumulative COVID-19 incidence through the cutoff date of September 19, 2020, where the first quartile has the highest cumulative incidence and the fourth quartile has the lowest incidence. Thus, the graph with the lightest colored data points shows the trend in gym visits by devices whose home was located in one of the lowest-incidence CSAs. By the week of April 6, gym visit rates among inhabitants of these lowest-incidence CSAs was 14.5 percent of baseline. By contrast, the darkest colored data points show that visit rates among inhabitants of the highest-incidence CSAs was 23.6 percent of baseline by that date.

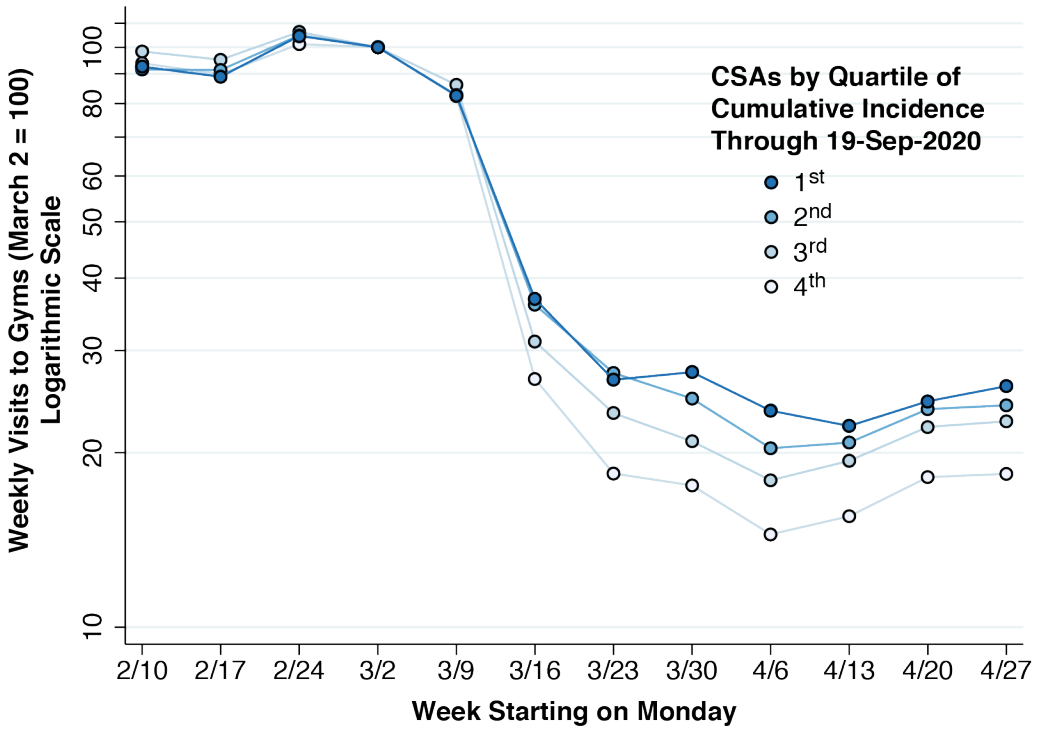


Figure 8. Weekly Visits to Gyms from Week of February 10 to Week of April 27, 2020, in Relation to the Quartile of Cumulative COVID-19 Incidence as of September 19, 2020

Figure 9 shows the corresponding trends in visits to gyms by month from February to August, 2020. In this case, the vertical axis has been normalized so that visits for the month of February equal 100. As in Figures 6 and 7, the time series corresponding to CSAs in the four quartiles of the logarithm of cumulative COVID-19 incidence as of September 19, 2020, where

the first quartile has the highest and the fourth quartile has the lowest incidence. During the month of May 2020, gym visits among residents of CSAs with the highest incidence was 29.9 percent of its February baseline, while gym visits among residents of CSAs with the lowest incidence was 23.9 percent of its February baseline. Thus, the gradient in gym visits across high- and low-incidence CSAs was largest in the month of April at the start of Phase II, but diminished progressively during Phases II and III.

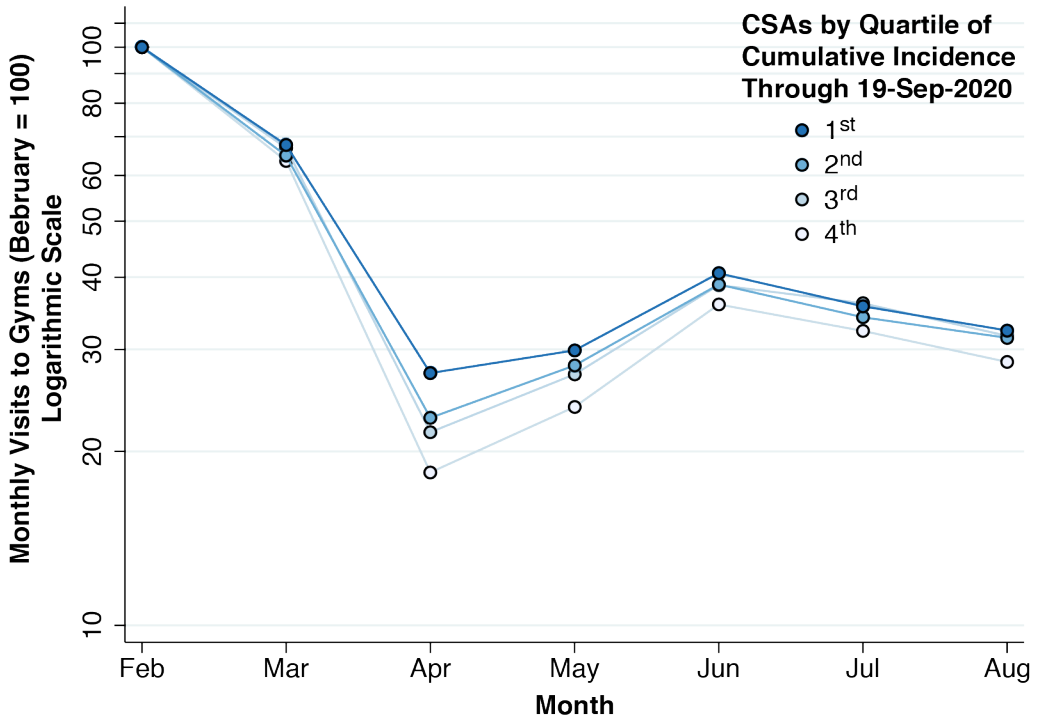


Figure 9. Monthly Visits to Gyms from February to August 2020, in Relation to the Quartile of Cumulative COVID-19 Incidence as of September 19, 2020

The lack of a complete series of weekly gym visitation rates prevented us from running Spatial Model 2 with gym attendance as an additional covariate. We were, however, able to carry out an analysis comparable to that for Figure 7. To that end, Figure 10 below explores the quantitative relationship between relative gym visitation during the month of April and the number of newly confirmed COVID-19 infections during the subsequent month of May. As in Figure 7, each data point is a single CSA. As indicated in Appendix D, the estimated population-weighted least squares regression line had a slope of 0.018, implying that a 10-percentage-point increase in gym visitation was associated with an 18-percent increase in newly confirmed cases.

These findings are further consistent with an impact of reduced social mobility on COVID-19 incidence during Phase II of the epidemic.

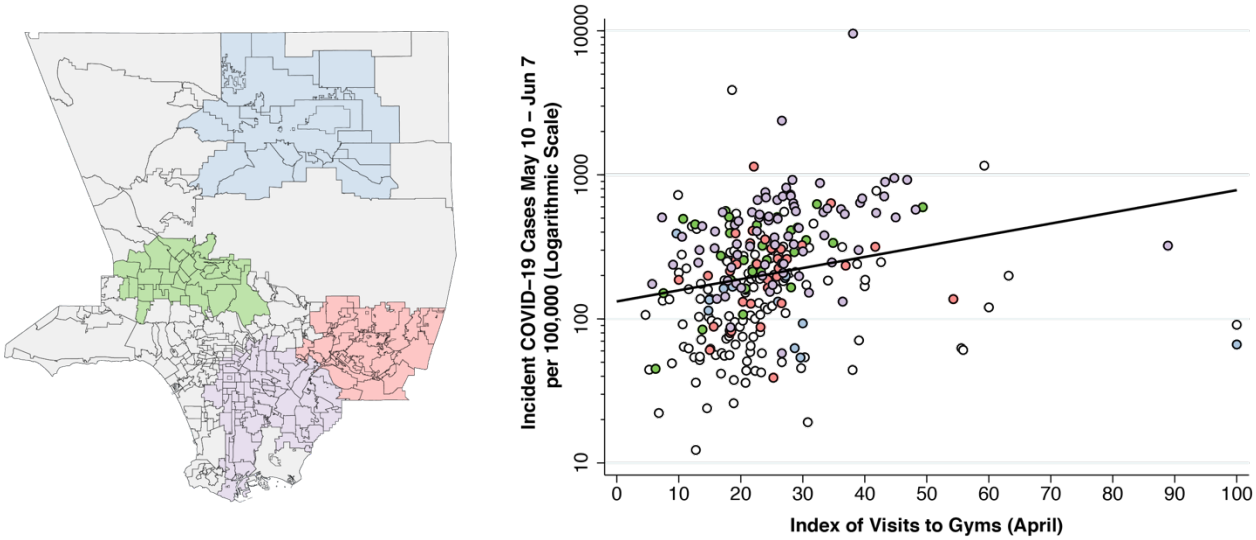


Figure 10. Relation Between Index of Visits to Gyms During April and the Incidence of Newly Confirmed Cases During May 10 – June 7, 2020

Household Structure

Figure 11 below shows the weekly incidence per 100,000 of newly confirmed COVID-19 cases from the week ending March 28 to the week ending September 19, 2020. We show four time series corresponding to CSAs in the four quartiles of at-risk household prevalence. The first quartile had an average prevalence of 25 percent, the second quartile had an average prevalence of 17 percent, the third had an average prevalence of 12 percent and the fourth had an average prevalence of 5 percent. As a result of data limitations described above, the time axis in Figure 11 omits the first two weeks of data available only at the countywide level, as shown in Figure 1.

During Phase I, weekly COVID-19 incidence was significantly higher among those CSAs with the *lowest* prevalence of at-risk households. During Phases II and III, the relationship reversed. By the week ending May 30, COVID-19 incidence among CSAs in the highest quartile was 4-fold the incidence among CSAs in the lowest quartile. This relative risk remained at 3 or more through the week ending August 15, after which it gradually decreased to 2.2 by the last week of our observation period.

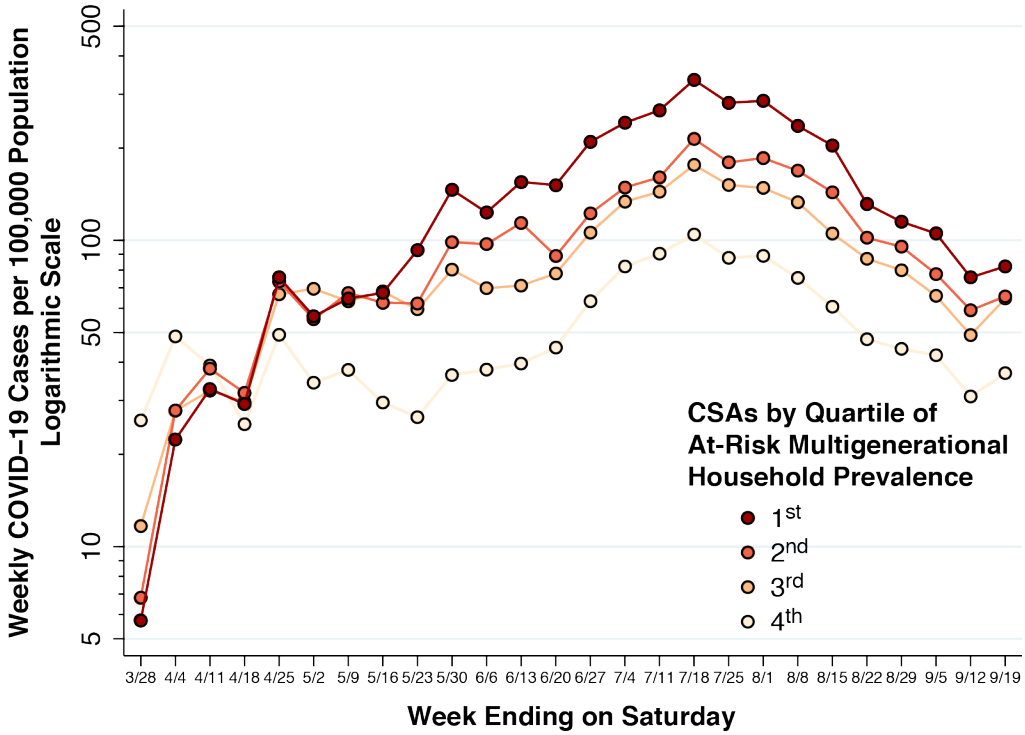


Figure 11. Weekly Incidence of Newly Confirmed COVID-19 Cases from the Week Ending March 28 to the Week Ending September 19, in Relation to the Quartile of Prevalence of At-Risk Multigenerational Households

Table B3 in Appendix B shows the results of Spatial Model 2 with two covariates: the percentage of devices staying completely at home (δ_1) and the percentage of at-risk multigenerational families (δ_2). During Phase I, we observed the same reverse relationships seen in Figures 6 and 11. As in Table B2, the percentage of devices staying completely at home was associated with a significant reduction in weekly confirmed COVID-19 cases. By contrast, the percentage of at-risk households was associated with a significant increase in COVID-19 cases not only in Phase II, but also in Phase III.

We have thus far relied upon raw counts of confirmed COVID-19 cases without reference to their age distribution. One concern is that households satisfying our at-risk criterion necessarily have at least one person over 50 years of age, and therefore CSAs with a higher prevalence of at-risk household also have more elderly inhabitants. Older individuals, in turn, may be more likely to experience symptoms, undergo testing, and thus be included in reported COVID-19 case counts. To address this concern, Figure 12 below relates the prevalence of at-risk households to the age-adjusted cumulative case rate as of the cutoff date of September 19, which is also reported by the DPH (Los Angeles County Department of Public Health 2020a).

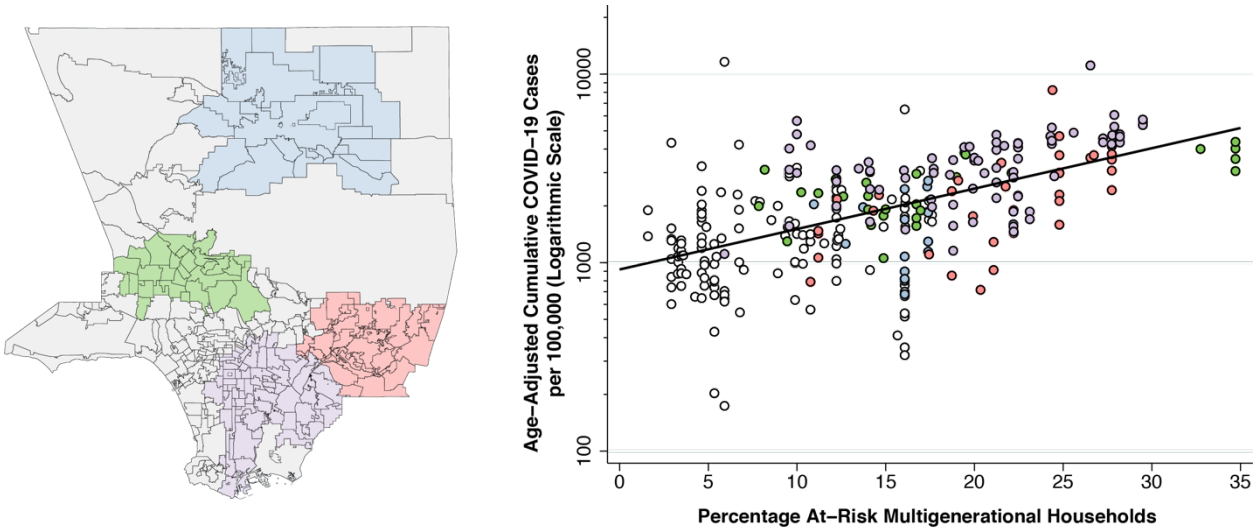


Figure 12. Relation Between Percentage of At-Risk Multigenerational Households and the Age-Adjusted Cumulative Incidence of Confirmed COVID-19 Cases During Through September 19, 2020

Figure 12 shows a positive relation between the prevalence of at-risk multigenerational households and age-adjusted cumulative incidence. As shown in Appendix D, the weighted least squares regression slope was 0.049, which means that every 10-percentage-point increase in the prevalence of at-risk households is associated with a 49-percent increase in cumulative incidence, with the population age distribution held constant.

Figure 13 below further explores the relationship between the prevalence of at-risk multigenerational households and the age-adjusted cumulative incidence as of September 19, 2020. To that end, we partitioned CSAs into four quartiles based upon their index of visits to gyms during April 2020, as described in Figures 9 and 10, and then tested the interaction between the index of gym visitation and the prevalence of at-risk households. As indicated in Appendix D, the weighted least squares regression line for the highest quartile of gym visitation, shown in green, had a slope of 0.052. The corresponding regression line for the lowest quartile of gym visitation, shown in gray, had a slope of 0.030. The difference in slopes, equal to 0.052 – 0.030 = 0.022, had a 95% confidence interval of [0.012 – 0.032] and was significantly different from zero ($p < 0.001$). These results support a significant interaction between gym visitation in April and at-risk household prevalence in determining accumulated COVID-19 cases by September 2020.

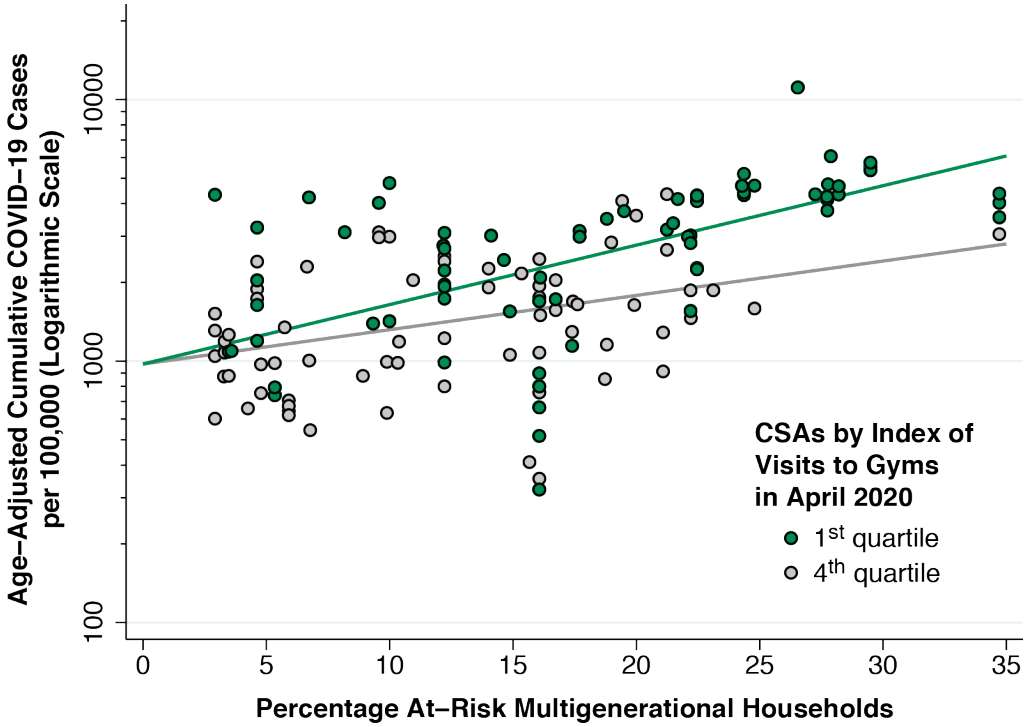


Figure 13. Relation Between Percentage of At-Risk Multigenerational Households and the Age-Adjusted Cumulative incidence of Confirmed COVID-19 Cases Through September 19, 2020 for CSAs in the Highest and Lowest Quartiles of the Index of Visits to Gyms in April 2020

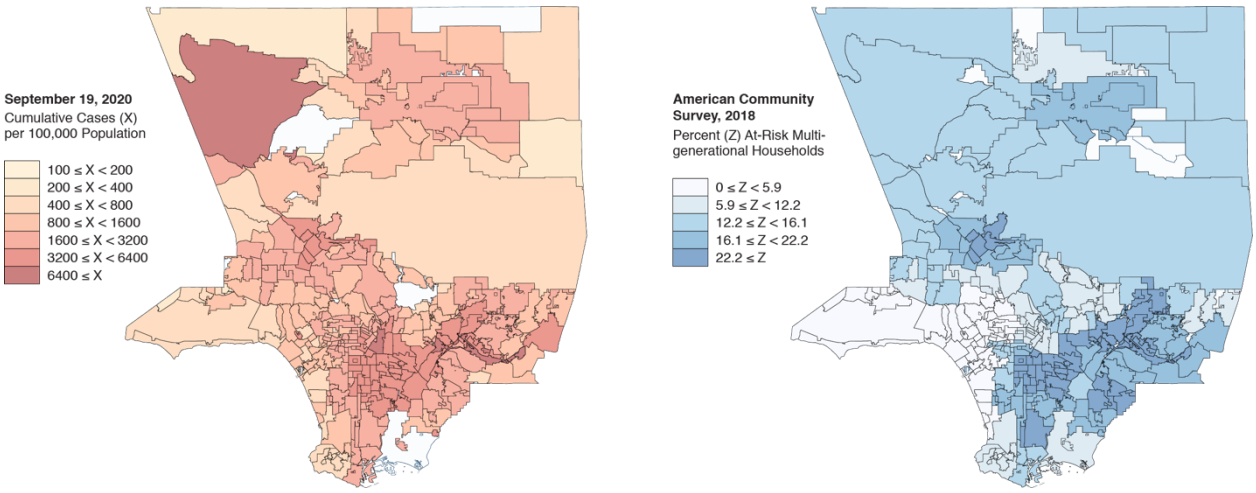


Figure 14. Cumulative Incidence of Confirmed COVID-19 Cases per 100,000 Population, Los Angeles County, as of September 11, 2020 (Left), Quintiles of the Percentage of At-Risk Multigenerational Families (Right)

Figure 14 above highlights the concordance between the prevalence of at-risk households and cumulative incidence with two maps of Los Angeles County. The map on the left repeats the geographic distribution of cumulative incidence through the September 19 cutoff date, shown

above in Figure 5. The map on the right shows the geographic distribution of the percentage of at-risk multigenerational households, color coded according to the quintile of at-risk prevalence. The four foci of COVID-19 infection identified in Figures 4 and 5 can be readily identified in the map on the right.

Discussion

Summary of Findings

We identified three phases of the COVID-19 epidemic in Los Angeles County (Figure 1). Phase I, which lasted until about April 4, saw the initial seeding of infection in relatively affluent communities such as Brentwood, Beverly Crest, and West Hollywood (Figure 2). From these initial foci, the epidemic spread radially (Figures 2 and 3). Beginning with Phase II, which saw COVID-19 incidence continue to rise until about July 11, the epidemic became concentrated in four identifiable foci (Figure 4) within the county. This geographic concentration was enhanced during Phase III (Figure 5), while COVID-19 incidence was falling.

These conclusions were supported not only qualitatively by examination of serial maps of cumulative COVID-19 incidence (Appendix A), but also quantitatively by estimates from a spatial adaption of the standard SIR epidemic model (Spatial Model 1, Appendix B). During Phase I, we found, the velocity of propagation of infection in adjoining community statistical areas (CSAs) within a 2-kilometer radius was an estimated 39 percent of the within-community velocity (Appendix Table B1). By Phase II, the estimated velocity of transmission to adjoining areas had fallen substantially, and by Phase III, it was indistinguishable from zero.

We used data on the movement of owners of smartphone devices equipped with location-tracking software to study two social mobility indicators: the percentage of device owners who stayed completely at home all day; and the number of device owners attending gyms. We found that both indicators responded significantly during Phase II of the epidemic, especially in the month of April (Figures 6, 8 and 9). Those CSAs with the largest reductions in social mobility in April exhibited the most significant attenuation of the epidemic in May (Figures 7 and 10, Appendix D).

We used data from the 2018 American Community Survey on the age structure and composition of households in each CSA to characterize the prevalence of households at risk for multigenerational spread of viral infection. Throughout both Phases II and III, the weekly

incidence of new COVID-19 infections in CSAs in the highest quartile of at-risk household prevalence was more than 3-fold the incidence in CSAs in the lowest quartile (Figure 11). The persistent relationship between the proportion of at-risk households and the weekly incidence of new infections was confirmed in a generalized spatial model (Spatial Model 2, Appendix B). The relationship between at-risk household prevalence and age-adjusted cumulative COVID-19 incidence was observed in all four foci of infection (Figure 12, Appendix D). The quantitative relation between at-risk household prevalence and age-adjusted cumulative COVID-19 incidence was strongest in those CSAs with the highest index of gym visitation in April (Figure 13, Appendix D). The map of the prevalence of at-risk multigenerational households is strikingly concordant with the map of cumulative incidence (Figure 13).

These findings, taken together, support the critical role of household structure in the persistent propagation of COVID-19 infections in Los Angeles County since April, once the epidemic had transitioned from its initial radial geographic spread to a phase of increasing concentration in high-incidence foci of infection. Moreover, we found evidence of a synergy between household structure and social mobility. The relation between the prevalence of at-risk multigenerational household and cumulative COVID-19 incidence was strongest in those CSAs with higher rates of gym attendance in April.

Strengths and Limitations of This Study

Our study takes advantage of the cohort structure of our database, in which we follow a group of related geographic units longitudinally over time. This structure allowed us to test a model of radial geographic expansion during Phase I of the Los Angeles County COVID-19 epidemic. It also allowed us to identify the month of April 2020 as the interval with the widest dispersion among CSAs in two indicators of social mobility, and then to test whether the levels of social mobility in April were predictive of the incidence of new infections in May. Such a research design overcomes many of the limitations of before-after studies (Harris 2020i).

On the other hand, our study is exclusively population-based. We do not follow a longitudinal cohort of individual households to see how many young adult members went to the gym, got infected, and then brought their infections home to older household members. A population-based indicator such as the proportion of households at risk for multigenerational transmission could thus be criticized as no more than a proxy for some other correlated characteristic of the community. One might inquire why we didn't test the relation between

cumulative COVID-19 incidence and the proportions of households relying on food stamps or speaking Spanish or headed by a single parent, to take but a few examples. These alternative covariates, however, are at best tangential to the key question of intrahousehold propagation of infection. If we had focused on the severity of the disease, and not simply the incidence, then population-based indicators of access to healthcare would be more relevant.

Our principal endpoint is the incidence of confirmed cases of COVID-19. It is now widely acknowledged that counts of confirmed cases substantially understate total infections (Havers et al. 2020). Serial population-based studies of seroprevalence are still uncommon (Hallal et al. 2020), and there is evidence that population seroprevalence may decline with time (Buss et al. 2020). Hospital admission rates have been studied as an alternative to confirmed case incidence (Harris 2020i, e), but such an endpoint depends on case severity and would thus be inappropriate for our study.

We relied upon two indicators of social mobility: the percentage of device holders staying completely at home; and the number of visitors to gyms. In other studies, we have measured attendance at bars (Harris 2020c), numbers of sit-down restaurant reservations (Harris 2020j, e) and indices of retail and recreational activity (Harris 2020d, e). These indicators are necessarily stand-ins for a more general concept. We cannot claim that attendance at gyms was the sole cause of the Los Angeles County epidemic any more than we can assert that attendance at sit-down indoor restaurants was the unique factor driving the surge of cases among younger persons in Florida (Harris 2020f). Still, we chose the rate of visitation at gyms as it would logically be an indicator of social activity among younger persons, and we likewise focused on the percentage staying completely at home as it would logically be an indicator of social immobility among older persons.

We noted that the exceptionally high cumulative incidence of COVID-19 infections in the unincorporated area of Castaic (Figure 4) was likely attributable to an outbreak at a local prison. Rather than attempt to identify and exclude other potential outliers in our universe of approximately 300 community statistical areas, we included all observations in our longitudinal analysis.

As a result of data limitations, our panel of CSA-specific case reports begins on March 21, 2020, well into Phase I of the Los Angeles County epidemic (Figure 1). The conclusion that the epidemic was already slowing down is supported by our estimate of a contemporaneous

reproductive number $\mathcal{R}$ = 2.0 for the interval March 21 – April 4, significantly below the basic reproductive number $\mathcal{R}_0$ = 3.5 observed in the earliest days of the outbreaks in Wuhan (Hao et al. 2020) and New York City (Harris 2020h).

While we don't have an accurate geospatial snapshot of the very earliest seeding of the virus, the fact that the cumulative incidence of confirmed cases had reached 100 per 100,000 population in relatively affluent CSAs (Figures 2 and 3) points to multiple importations by individuals with the resources to travel. A phylogenetic analysis of SARS-CoV-2 samples drawn during March 22 – April 15 at a major hospital located within one of the initial foci of infection found that the larger proportion belonged clades derived from Europe (Zhang et al. 2020). Our finding that the epidemic spread during Phase I principally through radial geographic expansion stands in sharp contrast to the earliest days of the outbreak in New York City, where community-transmitted infections were dispersed throughout all five boroughs in a matter of days (Harris 2020g, Gonzalez-Reiche et al. 2020).

To Live and Die from Coronavirus in L.A.

Having focused exclusively on the incidence of confirmed cases of COVID-19, this study has not confronted questions about coronavirus-related mortality. While we were able to reconstruct the temporal path of confirmed cases in the more than 300 communities in Los Angeles County, we did not have access to comparably detailed data on deaths by community by week. The study of mortality dynamics is considerably more complex, as it entails such additional considerations as access to medical care and the prevalence of co-morbidities that enhance the risk of complications. In Los Angeles County, in particular, the map of the COVID-19 case fatality rate may well turn out to look like a map of diabetes prevalence. Even further complicating the analysis is the progressive reduction in case fatality observed over the course of the epidemic (Harris 2020b).

Why Did Incidence Rates Undergo a Reversal in July 2020?

Thus far, we have made no mention of public policies concerning the COVID-19 epidemic, including orders issued by public officials at the state, county and municipal level. From the moment that the state governor (Newsom 2020), the county supervisor (Barger 2020), and the mayor of the city of Los Angeles (Garcetti 2020) all declared emergencies on March 4, these public policies undoubtedly contributed to the marked reduction in social mobility seen during Phase I and the beginning of Phase II (Figures 6, 8 and 9). The more difficult question is:

Did these public policies contribute to the turnaround in COVID-19 incidence that marked the transition from Phase II to Phase III at some time around week of July 12, 2020?

Figure 11 appears to show the near simultaneous peaking in COVID-19 incidence across diverse CSAs during the week of July 12-18. Appendix E, however, suggests that in the Antelope Valley and San Fernando Valley foci of infection, the peak incidence rate was delayed to the end of July. That would only widen the search window for the most influential public policies. On July 1, the Los Angeles County health officer issued an order closing indoor onsite dining (Los Angeles County Department of Public Health 2020d). On July 2, the office of the governor launched a wear-a-mask public awareness campaign (Office of Governor Gavin Newson 2020). On July 13, the state public health officer closed indoor operations in bars not concurrently serving meals, as well as gyms in counties on its monitoring list, to which Los Angeles County already belonged (Angell 2020). And on July 23, the Los Angeles County health officer ordered everyone diagnosed with or likely to have COVID-19 to self-isolate and quarantine (Davis 2020a, b).

Whether any one of these specific public policy measures had a dominant impact is COVID-19 incidence is difficult to ascertain. Figure 6 shows only a slight, temporary bump in the proportion of devices staying completely at home in late July and early August, while Figure 9 indicates that gym visitation in Los Angeles County in August 2020 was still hovering at 30 percent of its baseline level in February. The present study did not measure the prevalence of mask usage in Los Angeles County.

Implications

Despite an array of aggressive public policies aimed at reducing social mobility, our findings suggest that intrahousehold transmission has been the critical vehicle for the persistence of the COVID-19 epidemic in Los Angeles County. The prevalence of at-risk households in a community, it appears, is not simply a predictor of the persistence of coronavirus transmission, but also a multiplier of the effects of other policies aimed at social distancing. The impact of preventing one case of asymptomatic infection in a socially active young adult, who would otherwise have brought his or her infection into the household, will depend directly on the number of susceptible household members who have been spared.

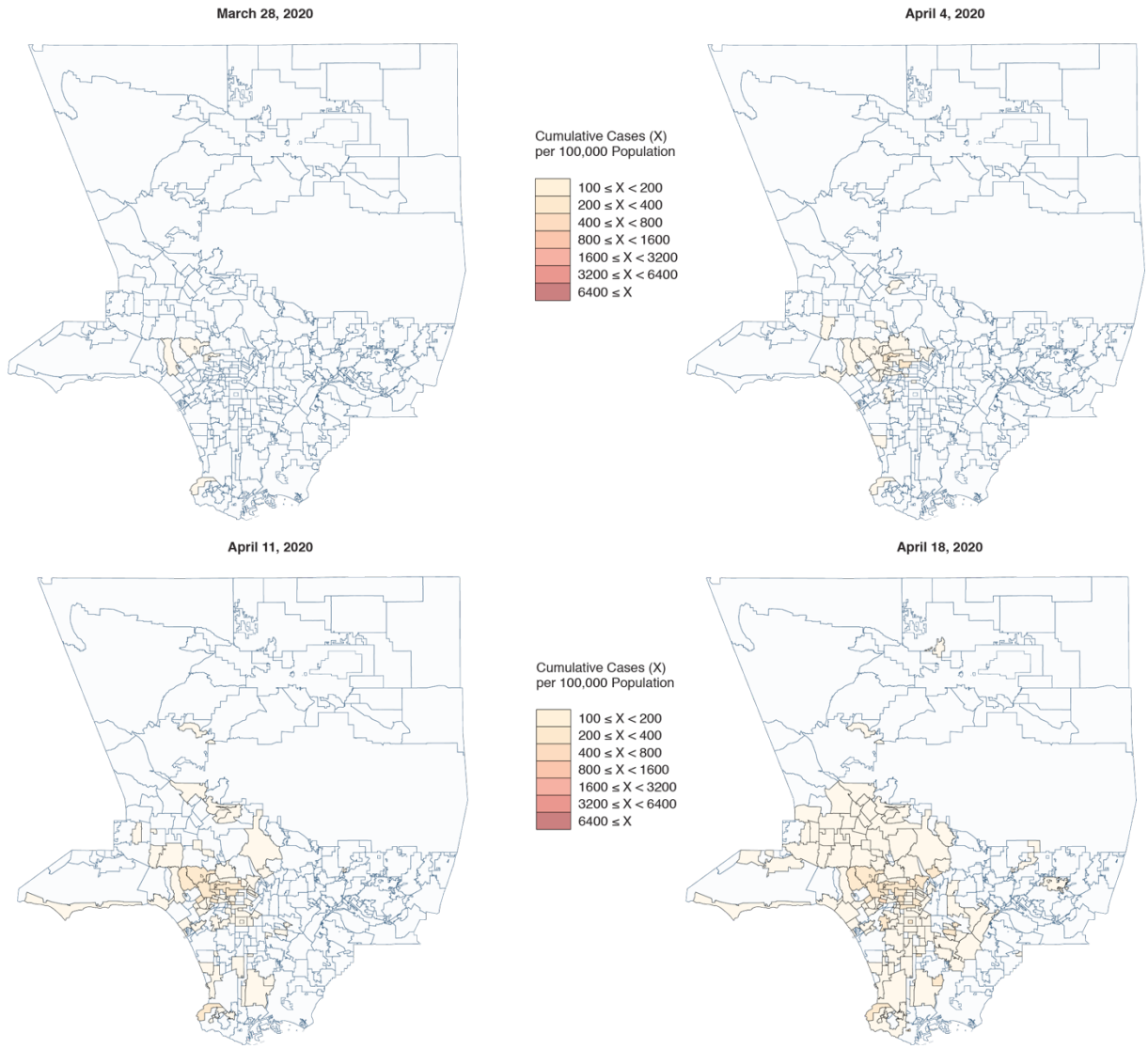
Our results cast a pessimistic shadow on so-called targeted policies that selectively relax restrictions on lower-risk, younger persons while seeking to protect more vulnerable older

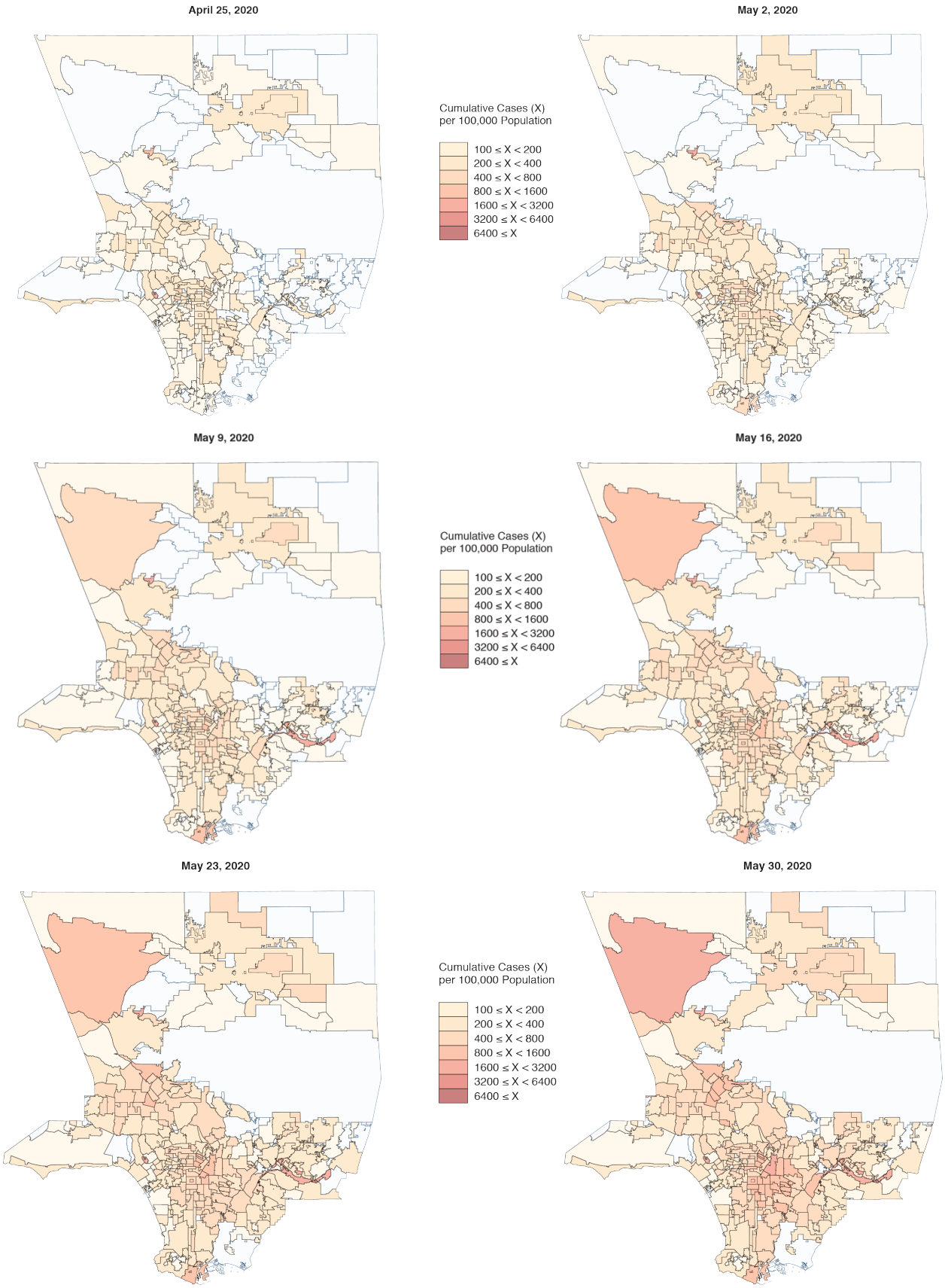
persons (Chikina and Pegden 2020, Acemoglu et al. 2020, Iverson, Karp, and Peri 2020, Gollier 2020). Such a policy might be feasible in settings where older persons are sequestered in retirement communities or assisted living facilities, but the data here show that this is not the reality of Los Angeles County.

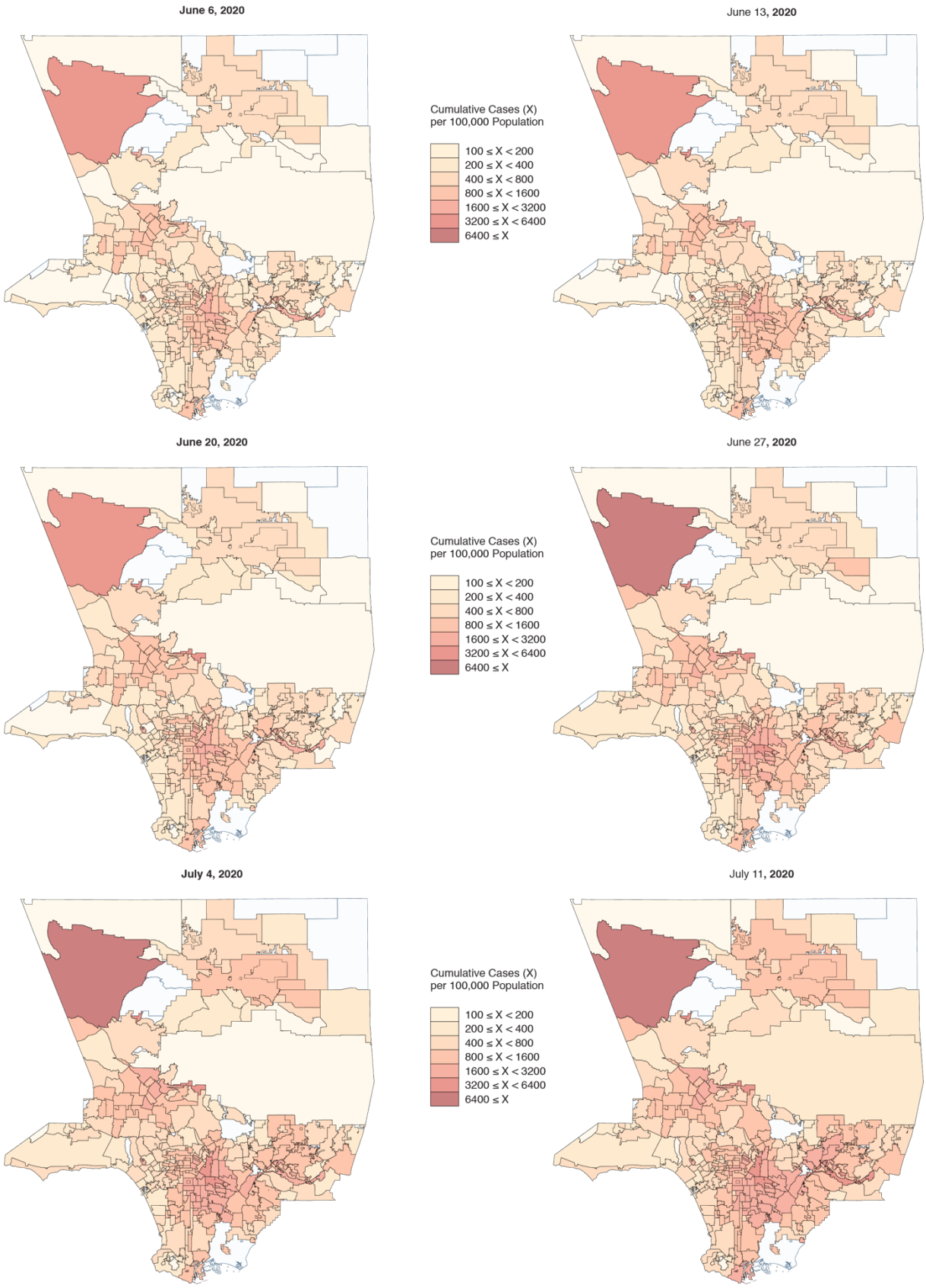
Most importantly, our findings require us to view the household rather than the individual as the foremost target of healthcare policy. The message "protect yourself" (*protégete* in Spanish) needs to be reconfigured as "protect your family" (*protége a tu familia*). When a healthcare provider encounters a new patient with suspected or established COVID-19, the interview needs to turn quickly to questions about other household members, their health status, and their symptoms. The widely recognized model of the patient-centered medical home (Alexander and Bae 2012) needs to be replaced by the family- and household-centered medical home.

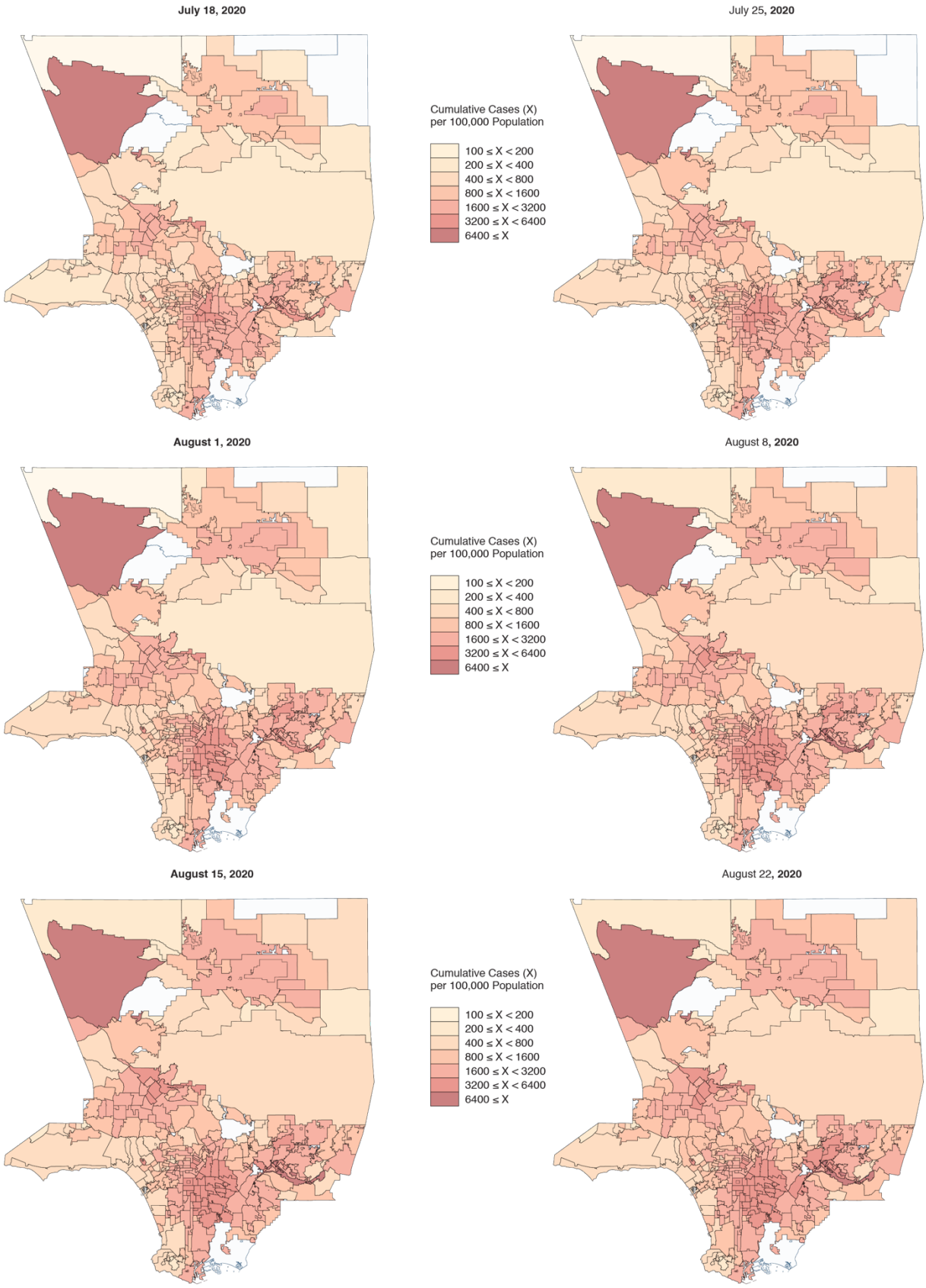
Appendix A: Cumulative Incidence of Confirmed Covid-19 Cases Per 100,000 Population, Los Angeles County, March 28 – September 19, 2020

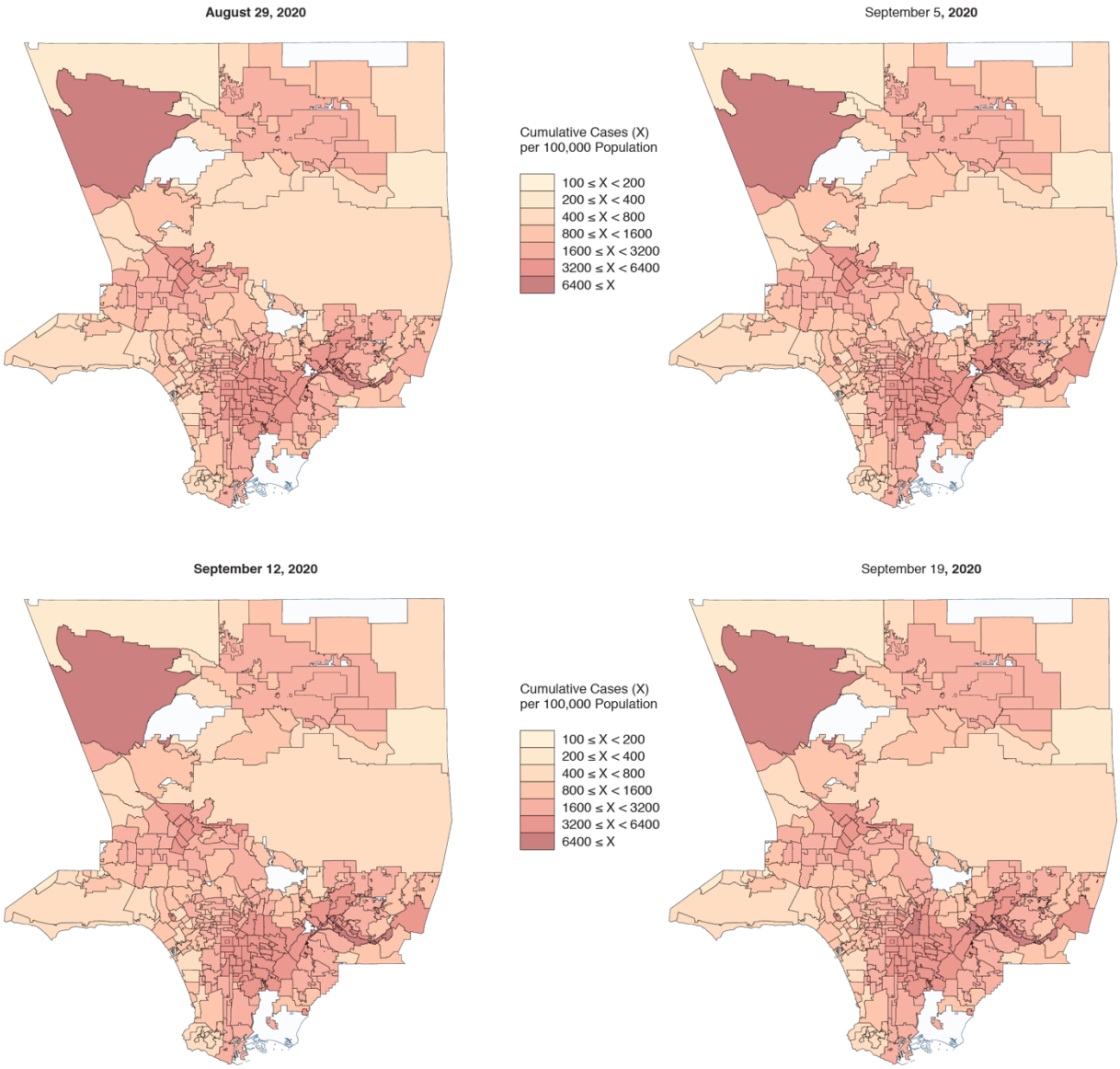
The maps shown below indicate the cumulative incidence per 100,000 in each countywide statistical area (CSA) for 26 successive weeks during March 28 – September 19, 2020. The DPH did not report COVID-19 counts for two cities within the county, Pasadena and Long Beach, which are colored pale blue throughout. CSAs are color-coded in a geometric progression, starting with a threshold of 100 cases per 100,000, and then successive doubling to 6,400 cases per 100,000. In the map for March 21, omitted here, no CSA had yet had a cumulative incidence over the 100-per-100,000 threshold.




April 25, 2020
May 2, 2020
May 9, 2020
May 16, 2020
May 23, 2020
May 30, 2020
Cumulative Cases (X)
per 100,000 Population
100 ≤ X < 200
200 ≤ X < 400
400 ≤ X < 800
800 ≤ X < 1600
1600 ≤ X < 3200
3200 ≤ X < 6400
6400 ≤ X


June 6, 2020
June 13, 2020
June 20, 2020
June 27, 2020
July 4, 2020
July 11, 2020
Cumulative Cases (X)
per 100,000 Population
100 ≤ X < 200
200 ≤ X < 400
400 ≤ X < 800
800 ≤ X < 1600
1600 ≤ X < 3200
3200 ≤ X < 6400
6400 ≤ X


July 18, 2020
July 25, 2020
August 1, 2020
August 8, 2020
August 15, 2020
August 22, 2020
Cumulative Cases (X)
per 100,000 Population
100 ≤ X < 200
200 ≤ X < 400
400 ≤ X < 800
800 ≤ X < 1600
1600 ≤ X < 3200
3200 ≤ X < 6400
6400 ≤ X


August 29, 2020
September 5, 2020
Cumulative Cases (X)
per 100,000 Population
100 ≤ X < 200
200 ≤ X < 400
400 ≤ X < 800
800 ≤ X < 1600
1600 ≤ X < 3200
3200 ≤ X < 6400
6400 ≤ X
September 12, 2020
September 19, 2020
Cumulative Cases (X)
per 100,000 Population
100 ≤ X < 200
200 ≤ X < 400
400 ≤ X < 800
800 ≤ X < 1600
1600 ≤ X < 3200
3200 ≤ X < 6400
6400 ≤ X

Appendix B: Detailed Spatial Model Estimates

In the detailed results described below, Phase I covers the period from March 21 – April 4, Phase II covers April 11 – July 11, and Phase III covers July 18 – September 19.

Table B1 shows the parameter estimates for Spatial Model 1 under the restriction that the constant term equals zero. As noted in equation (1) in the main text, the parameter α measures the effect of the within-CSA stock of infective individuals, while the parameter γ measures the effect of the stock of infectives in nearby CSA's within a 2-kilometer radius.

Table B1. Estimates of the Parameters of Spatial Model 1 for the Three Phases of the Los Angeles County Epidemic

Phase	$\hat{\alpha}$	$\hat{\gamma}$	No. Obs.
I	1.028	0.406	586
	(0.068)	(0.054)	
II	0.554	0.069	4,102
	(0.011)	(0.007)	
III	0.578	0.027	2,930
	(0.028)	(0.018)	

Table B2 shows the parameter estimates for Spatial Model 2 with an unrestricted constant term, where the parameter δ gauges the effect of the percentage of devices staying completely at home.

Table B2. Estimates of the Parameters of Spatial Model 2 for the Three Phases of the Los Angeles County Epidemic: Percent of Devices Staying Completely at Home

Phase	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\delta}$	No. Obs.
I	–16.403	0.724	0.228	0.643	584
	(7.469)	(0.068)	(0.050)	(0.167)	
II	118.919	0.471	0.027	–2.075	4,088
	(10.066)	(0.012)	(0.006)	(0.231)	
III	–2.565	0.493	0.013	0.886	2,922
	(37.016)	(0.038)	(0.018)	(0.983)	

Table B3 shows the parameter estimates for Spatial Model 2 with an unrestricted constant term, where parameters δ_1 and δ_2, respectively, gauge the effects of the percentage of devices

staying completely at home and the percentage of at-risk multigenerational households. The numbers of observations were identical to those in Table 2.

Table B3. Estimates of the Parameters of Spatial Model 2 for the Three Phases of the Los Angeles County Epidemic: (1) Percent of Devices Staying Completely at Home, (2) Percent of At-Risk Multigenerational Households

Phase	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\delta}_1$	$\hat{\delta}_2$
I	–7.500	0.649	0.228	0.613	–0.464
	(7.811)	(0.131)	(0.051)	(0.166)	(0.131)
II	87.420	0.445	0.034	–1.884	1.709
	(10.494)	(0.012)	(0.006)	(0.230)	(0.180)
III	–37.568	0.431	0.020	0.911	2.960
	(38.237)	(0.041)	(0.018)	(0.981)	(0.834)

We reran the spatial models shown in Tables B1–B3 with variations in the radius of inter-community influence (baseline: 2 kilometers), the mean serial interval (baseline: 5.5 days), and the definition of an at-risk household (baseline: one person 18–34 years, another person 50+ years, at least 4 persons). The results (not shown) displayed the same qualitative patterns seen here.

Appendix C: Interquartile Differences in the Percent Staying Completely at Home

To quantitatively test the relationships shown in Figure 6, we calculated the difference between the highest- and lowest-incidence quartiles in the logarithm of percent staying completely at home. The table below shows the means and standard errors of these logarithmic differences in relationship to the phase of the epidemic. While Phase I of the epidemic in fact began at the beginning of March, we have reclassified the weeks of March 1 and March 8 in the baseline period for the purposes of this comparison, inasmuch as public awareness of the threat of contagion did not begin until mid-March. The table shows that during the baseline period, the percent staying completely at home in the highest quartile of COVID-19 incidence was 22.4% *less* than the percent staying completely home in the lowest quartile of COVID-19 incidence. During phases I and II, the percent staying at home in the highest quartile was 5–6% *more* than the corresponding percent staying completely at home in the lowest quartile. By phase III, this gradient persisted, but was much smaller in magnitude.

Difference Between the Highest- and Lowest Incidence CSA Quartiles in the Logarithm of the Percent of Devices Staying at Home

Phase	Weeks	Mean	St. Error
Baseline	2/2 – 3/8	–0.224	0.020
I	3/15 – 4/5	0.051	0.009
II	4/12 – 7/5	0.059	0.012
III	7/12 – 9/13	0.019	0.005

Appendix D: Regression Models Underlying Figures 7, 10, 12 and 13

For Figure 7, we ran a weighted least squares regression across CSAs, where the dependent variable was the logarithm of the incidence of newly confirmed COVID-19 cases during May 10 – June 7, and the independent variable was the percentage of devices staying completely at home during April 5 – May 2. The regression weights were the logarithms of the CSA population. Standard errors are shown in parentheses under the parameter estimates.

Table D1. Weighted Least Squares Regression for Figure 7

Slope	Intercept	No. Obs.	R^2 statistic
–0.073	8.739	277	0.122
(0.012)	(0.555)		

For Figure 10, we ran a weighted least squares regression across CSAs, where the dependent variable was the logarithm of the incidence of newly confirmed COVID-19 cases during May 10 – June 7, and the independent variable was the index of visits to gyms during April 2020. As in Figure 7, the regression weights were the logarithms of the CSA population.

Table D2. Weighted Least Squares Regression for Figure 10

Slope	Intercept	No. Obs.	R^2 statistic
0.018	4.884	277	0.051
(0.005)	(0.125)		

For Figure 12, we ran a weighted least squares regression across CSAs, where the dependent variable was the logarithm of the age-adjusted cumulative incidence of confirmed COVID-19 cases through September 19, 2020, and the independent variable was the percentage of at-risk multigenerational households. As in Figures 7 and 20, the regression weights were the logarithms of the CSA population. Standard errors are again shown in parentheses.

Table D3. Weighted Least Squares Regression for Figure 12

Slope	Intercept	No. Obs.	R^2 statistic
0.049	6.823	293	0.367
(0.004)	(0.064)		

For Figure 13, we again ran a weighted least squares regression across CSAs, where the dependent variable was the logarithm of the age-adjusted cumulative incidence of confirmed COVID-19 cases through September 19, 2020. The independent variables included not only the percentage of at-risk multigenerational households, but also interaction terms with binary indicators of the quartile of gym visitation in April 2020. In the table below, the 1st quartile represents those CSAs with the highest index of gym visits (> 27.9 percent of the February 2020 volume), the 2nd quartile represents CSAs with an index of gym visits in the range of 22.7 – 27.9 percent of the February 2020 volume, and the 3rd quartile represents CSAs with an index of gym visits in the range of 18.0 – 22.7 percent of the February 2020 volume. The interaction terms with the binary indicator for the 4th quartile, representing CSAs with an index of gym visits less than 18.0 percent of the February 2020 volume, was the omitted variable.

Table D4. Weighted Least Squares Regression for Figure 13

Regressor	Parameter	St. Error
Percent At-Risk Households	0.030	(0.006)
× 1st Quartile Gym Visits	0.022	(0.005)
× 2nd Quartile Gym Visits	0.019	(0.005)
× 3rd Quartile Gym Visits	0.018	(0.005)
Intercept	6.881	(0.061)
No. Observations	286	
R^2 statistic	0.432	

In Figure 13, the gray regression line, corresponding to the lowest quartile of gym visitation, has a slope of 0.030, while the green regression line, corresponding to the highest quartile of gym visitation, has a slope of 0.030 + 0.022 = 0.052. The difference in slopes, estimated to be 0.022, had a 95% confidence interval of [0.012 – 0.032], and a t-test of the null hypothesis of no interaction yielded $p < 0.001$.

We reran the regression models shown in Tables D1–D4 with variations in the definition of an at-risk household (baseline: one person 18–34 years, another person 50+ years, at least 4 persons). The results (not shown) displayed the same qualitative patterns seen here.

Appendix E: Weekly Incidence of Newly Confirmed COVID-19 Cases by Focus of Infection

Figure E1 graphs the weekly incidence of newly confirmed COVID-19 cases for each of the four foci of infection. The unfilled data points refer to the remaining CSAs not located in any of the foci.

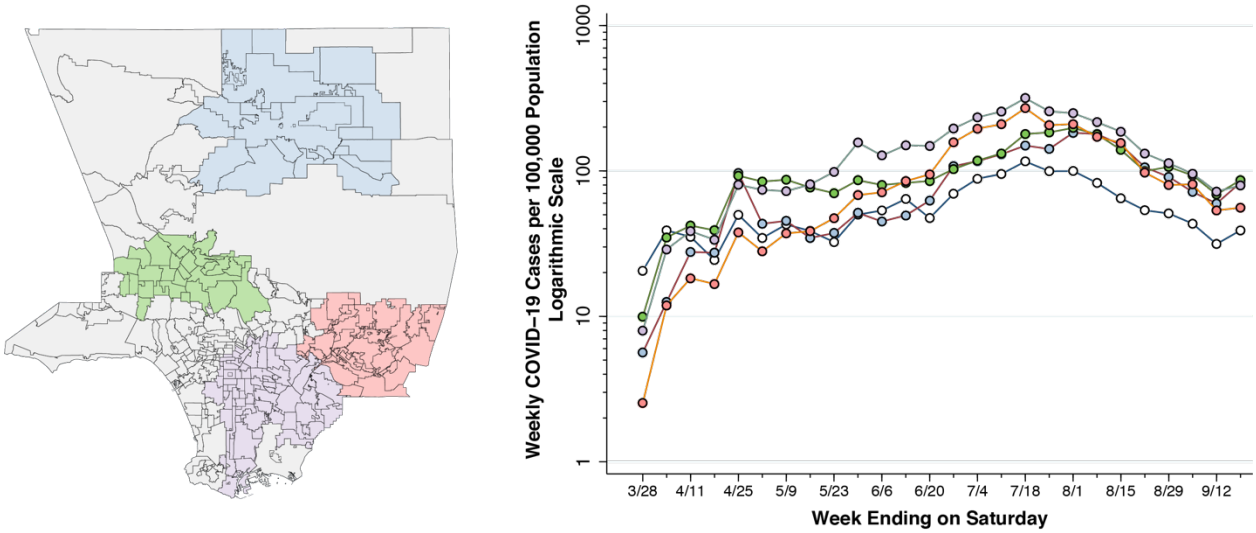


Figure E1. Weekly Incidence of Newly Confirmed COVID-19 Cases by Focus of Infection

References

Acemoglu, Daron, Victor Chernozhukov, Ivan Werning, and Michael D. Whinston. 2020. *Optimal Targeted Lockdowns in a Multi-Group SIR Model*. https://www.nber.org/papers/w27102: National Bureau of Economic Research Working Paper 27102, June 2020.

Alexander, J. A., and D. Bae. 2012. "Does the patient-centred medical home work? A critical synthesis of research on patient-centred medical homes and patient-related outcomes." *Health Serv Manage Res* 25 (2):51-9. doi: 10.1258/hsmr.2012.012001.

Angell, Sonia Y. 2020. *Statewide Public Health Officer Order*. https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/COVID-19/SHO%20Order%20Dimming%20Entire%20State%207-13-2020.pdf: California Department of Public Health, July 13, 2020.

Aparicio, Ainoa, and Shoshana Grossbard. 2020. *Intergenerational Residence Patterns and Covid-19 Fatalities in the EU and the US*. University of Turin, Italy.: September 9, 2020.

Barger, Kathryn. 2020. *Proclamation of Existence of a Local Emergency by the Chair of the Los Angeles County Board of Supervisors and Request for a State Proclamation*. https://kathrynbarger.lacounty.gov/wp-content/uploads/2020/03/SKMBT_C65220030412030.pdf: Los Angeles, March 4, 2020.

Buss, Lewis F., Carlos Augusto Jr. Prete, Claudia MM Abrahim, and et al. 2020. *COVID-19 herd immunity in the Brazilian Amazon*. https://www.medrxiv.org/content/10.1101/2020.09.16.20194787v1: medRxiv September 21, 2020.

Chikina, M., and W. Pegden. 2020. "Modeling strict age-targeted mitigation strategies for COVID-19." *PLoS One* 15 (7):e0236237. doi: 10.1371/journal.pone.0236237.

City of Los Angeles. 2020. *Countywide Statistical Areas (CSA)*. https://geohub.lacity.org/datasets/lacounty::countywide-statistical-areas-csa: September 19, 2020.

Dave, Dhaval, Andrew I. Friedson, D. McNichols, and Joseph J. Sabia. 2020. *The Contagion Externality of a Superspreading Event:The Sturgis Motorcycle Rally and COVID-19*. http://ftp.iza.org/dp13670.pdf: IZA Institute of Labor Economics, Discussion Paper 1360, September 2020.

Davis, Muntu. 2020a. *Health Officer Order for the Control of COVID-19: Public Health Emergency Isolation Order*. http://publichealth.lacounty.gov/media/Coronavirus/docs/HOO/HOO_Coronavirus_Blanket_Isolation.pdf: Los Angeles County Department of Public Health, Issued July 23.

Davis, Muntu. 2020b. *Health Officer Order for the Control of COVID-19: Public Health Emergency Quarantine Order*. http://ph.lacounty.gov/media/Coronavirus/docs/HOO/HOO_Coronavirus_Blanket_Quarantine.pdf: Los Angeles County Department of Public Health, Issued July 23.

Dickson, M. M., G. Espa, D. Giuliani, F. Santi, and L. Savadori. 2020. "Assessing the effect of containment measures on the spatio-temporal dynamic of COVID-19 in Italy." *Nonlinear Dyn*:1-14. doi: 10.1007/s11071-020-05853-7.

Esteve, A., I. Permanyer, D. Boertien, and J. W. Vaupel. 2020. "National age and coresidence patterns shape COVID-19 vulnerability." *Proc Natl Acad Sci U S A* 117 (28):16118-16120. doi: 10.1073/pnas.2008764117.

Fang, Y., Y. Nie, and M. Penny. 2020. "Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis." *J Med Virol* 92 (6):645-659. doi: 10.1002/jmv.25750.

Franch-Pardo, I., B. M. Napoletano, F. Rosete-Verges, and L. Billa. 2020. "Spatial analysis and GIS in the study of COVID-19. A review." *Sci Total Environ* 739:140033. doi: 10.1016/j.scitotenv.2020.140033.

Garcetti, Eric. 2020. *Declaration of Local Emergency*. http://clkrep.lacity.org/onlinedocs/2020/20-0291_reso_03-04-2020.pdf: Los Angeles, March 4, 2020.

Gollier, Christian. 2020. *Cost-benefit analysis of age-specific deconfinement strategies*. https://www.institutlouisbachelier.org/wp-content/uploads/2020/05/cost-benefit-analysis-of-age-specific-deconfinement-strategies-christian-gollier-tse.pdf: Toulouse School of Economics, May 25, 2020.

Gonzalez-Reiche, A. S., M. M. Hernandez, M. J. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Alburquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y. C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A.

Garcia-Sastre, M. Luksza, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon, and H. van Bakel. 2020. "Introductions and early spread of SARS-CoV-2 in the New York City area." *Science* 369 (6501):297-301. doi: 10.1126/science.abc1917.

Griffin, John M., Aine B. Collins, Kevin Hunt, and et al. 2020. *A rapid review of available evidence on the serial interval and generation time of COVID-19*. https://www.medrxiv.org/content/10.1101/2020.05.08.20095075v1: MedRxiv May 11, 2020.

Hallal, P. C., F. P. Hartwig, B. L. Horta, M. F. Silveira, C. J. Struchiner, L. P. Vidaletti, N. A. Neumann, L. C. Pellanda, O. A. Dellagostin, M. N. Burattini, G. D. Victora, A. M. B. Menezes, F. C. Barros, A. J. D. Barros, and C. G. Victora. 2020. "SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys." *Lancet Glob Health*. doi: 10.1016/S2214-109X(20)30387-9.

Hao, X., S. Cheng, D. Wu, T. Wu, X. Lin, and C. Wang. 2020. "Reconstruction of the full transmission dynamics of COVID-19 in Wuhan." *Nature* 584 (7821):420-424. doi: 10.1038/s41586-020-2554-8.

Harris, Jeffrey E. 2020a. *The Coronavirus Epidemic Curve is Already Flattening in New York City*. https://www.nber.org/papers/w26917: National Bureau of Economic Research Working Paper No. 26917, Updated April 6, 2020.

Harris, Jeffrey E. 2020b. *COVID-19 Case Mortality Rates Continue to Decline in Florida*. https://www.medrxiv.org/content/10.1101/2020.08.03.20167338v1: medRxiv August 4, 2020.

Harris, Jeffrey E. 2020c. "COVID-19, bar crowding, and the Wisconsin Supreme Court: A non-linear tale of two counties." *Research in International Business and Finance* 54:101310. doi: doi.org/10.1016/j.ribaf.2020.101310.

Harris, Jeffrey E. 2020d. *Daily New COVID-19 Cases in Tulsa, Oklahoma through July 8, 2020*. https://lifeunderquarantine.blog/2020/07/09/daily-new-covid-19-cases-in-tulsa-oklahoma-through-july-8-2020/: Life Under Quarantine July 9, 2020.

Harris, Jeffrey E. 2020e. *Data From the COVID-19 Epidemic in Florida Suggest That Younger Cohorts Have Been Transmitting Their Infections to Less Socially Mobile Older Adults*. https://rdcu.be/b6pQu: Review of Economics of the Household doi: 10.1007/s11150-020-09496-w, August 2020.

Harris, Jeffrey E. 2020f. "Data from the COVID-19 epidemic in Florida suggest that younger cohorts have been transmitting their infections to less socially mobile older adults." *Review of Economics of the Household* (August 11, 2020). doi: doi.org/10.1007/s11150-020-09496-w.

Harris, Jeffrey E. 2020g. *The Earliest Days of the Coronavirus Outbreak in New York City: Part 1*. https://lifeunderquarantine.blog/2020/08/13/the-earliest-days-of-the-coronavirus-outbreak-in-new-york-city-part-1/: Life Under Quarantine, August 13, 2020.

Harris, Jeffrey E. 2020h. *The Earliest Days of the Coronavirus Outbreak in New York City: Part 2*. https://lifeunderquarantine.blog/2020/08/30/the-earliest-days-of-the-coronavirus-outbreak-in-new-york-city-part-2/: Life Under Quarantine, August 30, 2020.

Harris, Jeffrey E. 2020i. *Reopening Under COVID-19: What to Watch For*. http://web.mit.edu/jeffrey/harris/HarrisJE_WP3_COVID19_WWF_6-May-2020.pdf: May 12, 2020.

Harris, Jeffrey E. 2020j. *San Antonio Conundrum*. https://lifeunderquarantine.blog/2020/07/15/san-antonio-conundrum/: Life Under Quarantine, July 15, 2020.

Harris, Jeffrey E. 2020k. *TETRIS For Tulsa*. https://lifeunderquarantine.blog/2020/07/12/tetris-for-tulsa/: Life Under Quarantine, July 12, 2020.

Havers, Fiona P., Carrie Reed, Travis W. Lim, and et al. 2020. *Seroprevalence of Antibodies to SARS-CoV-2 in Six Sites in the United States, March 23-May 3, 2020*. https://www.medrxiv.org/content/10.1101/2020.06.25.20140384v1: MedRxiv June 26, 2020.

Hedges, Andrew. 2002. *Finding distances based on Latitude and Longitude*. https://andrew.hedges.name/experiments/haversine/: Accessed September 25, 2020.

Iverson, Terrence, Larry S. Karp, and Alessandro Peri. 2020. *Protecting the Vulnerable during a Pandemic with Uncertain Vaccine Arrival and Reinfection*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3648714: Social Science Research Network, July 12, 2020.

Los Angeles County Department of Public Health. 2020a. *COVID-19 Surveillance Dashboard*. https://lacdph.shinyapps.io/covid19_surveillance_dashboard/: Last accessed October 10, 2020.

Los Angeles County Department of Public Health. 2020b. *LA_County_Covid19_cases_deaths_date_table.csv*. http://dashboard.publichealth.lacounty.gov/covid19_surveillance_dashboard/: LA County COVID-19 Surveillance Dashboard, Accessed 20-Sep-2020.

Los Angeles County Department of Public Health. 2020c. *Listing of L.A. County D.P.H. Press Releases*. http://publichealth.lacounty.gov/phcommon/public/media/mediapubdisplay.cfm?unit=media&ou=ph&prog=media: Updated September 28, 2020.

Los Angeles County Department of Public Health. 2020d. *News Release: Revised Health Officer Order Requires Closure of Indoor On-Site Dining and Other Indoor Activities - 35 New Deaths and 2,002 New Cases of Confirmed COVID-19 in Los Angeles County*. http://publichealth.lacounty.gov/phcommon/public/media/mediapubhpdetail.cfm?prid=2477: July 1, 2020.

Newsom, Gavin. 2020. *Proclamation of a state of Emergency*. https://www.gov.ca.gov/wp-content/uploads/2020/03/3.4.20-Coronavirus-SOE-Proclamation.pdf: Executive Department, State of California, March 4, 2020.

Office of Governor Gavin Newson. 2020. *News Release: Governor Newsom Launches "Wear a Mask" Public Awareness Campaign in Response to Surge in COVID-19 Cases*. https://www.gov.ca.gov/2020/07/02/governor-newsom-launches-wear-a-mask-public-awareness-campaign-in-response-to-surge-in-covid-19-cases/: July 2, 2020.

Orea, Luis, and Imaculada C. Alvarez. 2020. *How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces* https://documentos.fedea.net/pubs/dt/2020/dt2020-03.pdf: FEDEA, April 13, 2020.

SafeGraph Inc. 2020a. *Places Schema*. https://docs.safegraph.com/docs/places-schema: Data accessed July 30-31, September 24-26, 2020.

SafeGraph Inc. 2020b. *Social Distancing Metrics*. https://docs.safegraph.com/docs/social-distancing-metrics: Accessed September 24, 2020.

Smith, Wyatt. 2020. *Pitchess Detention Center In Castaic Experiences COVID-19 Outbreak*. https://www.hometownstation.com/santa-clarita-news/coronavirus/pitchess-detention-center-in-castaic-experiences-covid-19-outbreak-326471: Santa Clarita News, May 7, 2020.

Tchekmedyian, Alene. 2020. *As officials ramp up testing in L.A. County jails, first inmate with COVID-19 dies* https://www.latimes.com/california/story/2020-06-01/correctional-health-inmates-testing-covid-coronavirus: Los Angeles Times, June 1, 2020.

U.S Census Bureau. 2018. *PUMS Technical Documentation: 2018*. https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html: Accessed September 19, 2020.

U.S Census Bureau. 2020a. *2010_Census_Tract_to_2010_PUMA.txt*. https://www2.census.gov/geo/docs/maps-data/data/rel/2010_Census_Tract_to_2010_PUMA.txt: Accessed September 13, 2020.

U.S Census Bureau. 2020b. *Public Use Microdata Areas (PUMAs)*. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html: August 18, 2020.

Zhang, W., J. P. Govindavari, B. D. Davis, S. S. Chen, J. T. Kim, J. Song, J. Lopategui, J. T. Plummer, and E. Vail. 2020. "Analysis of Genomic Characteristics and Transmission Routes of Patients With Confirmed SARS-CoV-2 in Southern California During the Early Stage of the US COVID-19 Pandemic." *JAMA Netw Open* 3 (10):e2024191. doi: 10.1001/jamanetworkopen.2020.24191.