Nursing home staff networks and COVID-19

M. Keith Chen[a,1,2] , Judith A. Chevalier[b,c,1], and Elisa F. Long[a,1]

[a]Anderson School of Management, University of California, Los Angeles, CA 90095; [b]Yale School of Management, Yale University, New Haven, CT 06511; and [c]National Bureau of Economic Research, Cambridge, MA 02138

Nursing homes and other long-term care facilities account for a disproportionate share of COVID-19 cases and fatalities worldwide. Outbreaks in US nursing homes have persisted despite nationwide visitor restrictions beginning in mid-March. An early report issued by the Centers for Disease Control and Prevention identified staff members working in multiple nursing homes as a likely source of spread from the Life Care Center in Kirkland, WA, to other skilled nursing facilities. The full extent of staff connections between nursing homes—and the role these connections serve in spreading a highly contagious respiratory infection—is currently unknown given the lack of centralized data on cross-facility employment. We perform a large-scale analysis of nursing home connections via shared staff and contractors using device-level geolocation data from 50 million smartphones, and find that 5.1% of smartphone users who visited a nursing home for at least 1 h also visited another facility during our 11-wk study period—even after visitor restrictions were imposed. We construct network measures of connectedness and estimate that nursing homes, on average, share connections with 7.1 other facilities. Traditional federal regulatory metrics of nursing home quality are unimportant in predicting outbreaks, consistent with recent research. Controlling for demographic and other factors, a home's staff network connections and its centrality within the greater network strongly predict COVID-19 cases.

nursing homes | COVID-19 | complex networks | smartphone data

Linked to more than 40% of all US fatalities as of August 31, 2020, nursing homes and other long-term care facilities have been disproportionately afflicted by the ongoing coronavirus pandemic (1–3).* With an elderly resident population, many with underlying chronic medical conditions, congregate living quarters, and routine contact with staff members and outside visitors, nursing homes are particularly vulnerable to outbreaks of respiratory pathogens (4, 5). The US Centers for Medicare and Medicaid Services (CMS), the primary federal regulator of nursing homes, estimates that more than 30% of all nursing home residents in New Jersey, Connecticut, and Massachusetts had contracted severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as of June 28, 2020, and that more than 9% of the entire nursing home population died in these states (6).

Evidence from the early outbreak at the Life Care Center in Kirkland, WA, demonstrated that nursing homes and other congregate facilities face extremely elevated risks of virus spread (7, 8). CMS guidance issued on March 13, 2020 significantly restricted visitor access to long-term care facilities—effectively locking down nursing homes to residents, staff, and contractors (9). Nevertheless, many COVID outbreaks subsequently occurred in nursing homes, suggesting the unwitting introduction of the virus into homes by staff and contractors as one potential channel. The practice of employing nursing home staff across multiple facilities may play a key role in the spread of SARS-CoV-2, as a US Centers for Disease Control and Prevention (CDC) report issued on March 18, 2020 identified staff working in multiple nursing homes as a likely source of spread from the Life Care Center to other skilled nursing facilities in Washington State (8). Of the first four nursing homes with COVID outbreaks following this initial outbreak, two facilities received

patient transfers from Life Care, and two facilities employed staff working in both places (10).

Despite this early recognition of cross-traffic between congregate settings as a potentially important transmission mode, the extent of connections between nursing homes remains unknown due to lack of systematic data. Furthermore, although the CDC identified staff members working in multiple long-term care facilities as a key high-risk group, CMS has not provided any specific guidance on this practice or on reducing contacts between homes more generally (9, 11, 12).

Using device-level geolocation data for 501,503 smartphones observed in at least one of the 15,307 nursing homes in the continental United States, we find that 5.1% of individuals who spent at least 1 h in a nursing home also spent at least 1 h in one or more other nursing homes in the 11-wk period following the March 13 nationwide restriction on nursing home visitors. We construct several measures from network theory to characterize nursing home connectedness, and examine whether such connectivity predicts confirmed and suspected COVID-19 cases. These data are anonymized, but, given the prohibition of social visitors, this cross-traffic between homes is likely traceable to staff and contractors. While our methodology cannot establish causation, we find that the number and strength of connections between nursing homes—and a home's centrality within the greater network—strongly predict COVID cases, even after controlling for location, demographic factors, number of beds, for-profit status, and CMS quality ratings. Consistent with recent

Significance

Nursing homes account for 40% of US COVID-related fatalities as of August 31, highlighting the urgent need to reduce SARS-CoV-2 transmission routes in these facilities. Our large-scale analysis of smartphone location data reports half a million individuals entering a nursing home following the March 13 federal ban on visitors. With 5.1% of these individuals entering two or more facilities, a nursing home snapshot network emerges. More connections, likely arising from contractors and staff working at multiple facilities, are highly predictive of COVID-19 cases, whereas traditional regulatory quality metrics are unimportant in predicting outbreak size. With an estimated 49% of nursing home cases attributable to cross-facility staff movement, attention to highly connected nursing facilities is warranted.

*As of September 24, 2020, the CMS documents 55,845 COVID-19 nursing home deaths, fewer than other cited sources. As discussed below, nursing home cases and deaths occurring before May 1, 2020 were not required to be reported to CMS.

research (13–15), we observe that traditional federal regulatory metrics of nursing home quality are unimportant in predicting outbreak size.

COVID-19 in Congregate Facilities

The high case count and death toll in long-term care facilities demonstrates the urgent need to understand how transmission mechanisms within these facilities are distinct from broader community spread, to guide targeted policy initiatives and testing strategies (16, 17). Given the incomplete case reporting by CMS, extant studies of nursing home cases typically rely upon researcher-compiled state data. Three studies (13–15) examine the relationship between cases, home location, home demographics, and CMS quality ratings for facilities in a number of states. No study finds CMS ratings to be significant explanators of cases, although demographics and urban location are predictive of cases. Two studies of individual states (18, 19) find that higher CMS-rated nursing homes report fewer cases. One analysis finds no evidence that for-profit status significantly predicts nursing home cases (14), yet a study of Connecticut facilities does find for-profit status to be a predictor of cases (20). While all of these papers provide careful statistical analysis of COVID in nursing home settings, no study directly measures connections among homes.

The importance of connections between congregate settings in SARS-CoV-2 spread has largely been identified through case studies rather than large-scale analysis. The CDC's evaluation of the Kirkland, WA, outbreak pointed specifically to staff employed at multiple nursing homes as a factor in spreading the initial outbreak to additional homes (8). A study of four nursing homes in London (21) finds that 11% of staff worked in multiple homes, and these workers were 3 times as likely to be infected as workers in a single home. Further, ref. 21 also shows that whole-genome sequencing of positive samples from residents and staff indicated cross-infection between residents and staff as well as multiple introductions of the virus into individual care homes. In a different congregate setting, movement of staff and residents across three affiliated homeless shelters likely contributed to outbreaks in each location (22). Employees at food processing plants are at increased risk of contracting SARS-CoV-2 given their proximate working conditions and frequent use of shared transportation between crowded, communal housing and the workplace (23).

The movement of incarcerated individuals and the cross-usage of staff across prisons have been identified as risk factors for COVID-19 outbreaks; incoming inmate transfers were the probable source of the San Quentin Prison outbreak (24, 25). While we focus on SARS-CoV-2, the importance of linkages between congregate settings has been identified in case studies of prior disease outbreaks. Each of the three flu outbreaks at San Quentin during the 1918 influenza pandemic were linked to the introduction of a single transferred prisoner from a facility where flu was prevalent (26).

In principle, if a congregate setting were completely closed to the outside, infection could not enter. A key challenge in isolating nursing homes derives from their reliance on staff who live in the community. A study by the state of New York (27) concluded, largely based on the timing of infections, that through no fault of their own, nursing home workers were likely the main source of SARS-CoV-2 transmission in nursing homes. They find that roughly one-quarter of nursing home workers in New York State tested positive for the virus. Below, we describe, briefly, nursing home staffing practices and how they may exacerbate disease spread.

Nursing Home Staffing Practices and Regulation

Even in nonpandemic times, nursing home staffing presents challenges. Resident census and health conditions fluctuate from day to day, altering staffing needs on a daily basis with unpredictable absences, complicating the staffing problem (28). Understaffing leads to poor service and regulatory violations, while overstaffing increases costs. To help manage this trade-off, care facilities often rely on staffing agencies to employ nurses and nurse aides and provide them on an on-call basis (28, 29). While data are limited, a 2009 study suggests that 60% of nursing homes use a staffing agency for some of their staffing (30). Given this widespread reliance on staffing agencies and the recent growth in nursing home chain affiliates (31), many nurses and nursing assistants commonly work in multiple facilities. Nursing homes also receive services from hospice workers, dialysis technicians, clinicians, medical transporters, and other nonnursing staff that visit multiple homes. In addition to this planned cross-usage, nursing home workers may combine employment across multiple nursing homes as well as other jobs. Survey data from 2012 indicate that 19% of nursing assistants and 13% of registered nurses hold a second job of some type (32). According to the Bureau of Labor Statistics, the median nursing assistant earned $28,980 in May 2019, which makes a willingness to work multiple jobs unsurprising. However, extant regulatory data at the nursing home level do not track the degree to which healthcare workers work in more than one nursing home or other healthcare setting.

Examination of the nursing home COVID-19 crisis is further hindered by the fact that CMS did not require nursing homes to submit data on COVID-19 cases and fatalities until May 2020. Thus, for our main data analysis, we use the disclosures of individual state Departments of Public Health to determine cumulative nursing home COVID cases. From the 22 states for which home-level resident case data are available, we collected data on cumulative resident cases as of May 31, 2020 (or closest reporting period). In *SI Appendix*, we repeat our analyses using the cumulative case data reported by CMS for homes nationwide, with the caveat that CMS instructions for reporting cumulative cases allowed nursing homes to not report cases occurring before May 2020. For example, the nation's first congregate COVID outbreak, the Life Care Center of Kirkland, WA, is recorded in CMS data as having a cumulative zero COVID-19 cases, while the CDC report (8) states that, as of March 18, 2020, 81 residents of the facility had contracted the virus, and 23 persons had died.

Using the CMS address of record for each facility, we merge the nursing home-level COVID-19 case data with nursing home staff network connections measured using anonymized device-level smartphone data for the continental United States over the 11-wk period March 13 to May 31, 2020. Summary statistics for the 22 states for which we assembled nursing home data and the set of US facilities regulated by CMS with complete data are given in Table 1.

Network Measures

Nursing homes display a wide range of connectedness with other homes. Average *degree*—the number of facilities a nursing home shares at least one smartphone connection with—across the United States is $\langle k \rangle = 6.6$, but ranges from an average degree below 1, in South Dakota, Vermont, and Wyoming, to an average exceeding 10, in Florida, Maryland, and New Jersey (*SI Appendix*, Table S1). Among nursing homes with confirmed or suspected cases reported to CMS, average degree is 7.8 compared to 5.6 among homes with no documented cases ($t = 16.1$, $p < 0.0001$), with a significant difference across the entire degree distribution (Fig. 1). Average *strength*—the total number of smartphones appearing in a home and its connections—is also greater in homes with COVID (9.7 vs. 7.5, $t = 7.9$, $p < 0.0001$). For example, the correlation between a home's degree in the first 5 wk of our data collection and the last 5 wk is 0.58; the correlation for strength is even higher, at 0.82.

Table 1. Summary statistics of US nursing homes

Variable	State reporting facilities	CMS reporting facilities
Number of nursing homes	6,337	13,165
Demographics		
High proportion (>25%) of Black residents, %	16.7	12.7
High proportion (>50%) on Medicaid, %	32.9	28.1
Urban location, %	81.2	72.5
Regulatory measures		
Number of beds	115 (59.1)	109 (60.3)
CMS quality rating (1–5)	3.18 (1.42)	3.15 (1.42)
Has infection violations, %	75.3	75.7
Network metrics		
Node degree	7.08 (8.38)	6.42 (7.89)
Node strength	8.82 (12.4)	8.11 (14.4)
Weighted average neighbor degree	10.21 (8.33)	9.42 (8.22)
Eigenvector centrality in state	0.095 (0.19)	0.087 (0.19)

CMS facilities include all continental US nursing homes that report demographic and regulatory data. Binary variables are percent of nursing homes; continuous variables are mean values, with standard deviations in parentheses.

To illustrate how network measures differ across nursing homes, we present network diagrams for a subset of homes in six states as depicted in Fig. 2 and summarized in Table 2. Nodes denote individual nursing homes, and edges represent connections between nodes (i.e., at least one smartphone observed in both homes). More-connected nodes are generally toward the center of each diagram, and nodes with fewer connections are on the periphery. In each subnetwork, a focal nursing home or "hub" is shown in blue, with its direct neighbors (homes with at least one shared contact) in dark gray and its neighbors' neighbors in light gray. Node size denotes CMS-reported confirmed and suspected COVID cases among residents as of May 31, 2020. Edge color corresponds to the number of unique smartphones observed in each pair of homes.

A major challenge facing nursing homes is that every connection is a potential link to other connections—and to SARS-CoV-2 transmission. In the Alabama subnetwork (Fig. 2A), for instance, the focal nursing home reported 8 COVID cases among residents and 30 confirmed or suspected cases among staff, and this facility is directly connected to another Alabama nursing home with 68 resident and 48 staff cases (the larger gray node). Both facilities are highly connected to other homes, including one nursing home that shared 43 smartphones with the focal home—after visitor restrictions were imposed in March. Although California nursing homes have average degree of 6.0 and average strength of 7.3, both slightly below US averages, one Los Angeles facility (Fig. 2B) has degree of 9 and strength of 83, implying that homes connected to this hub share, on average, nine staff members, each of whom may be a potential conduit of SARS-CoV-2 transmission given the home's 63 reported cases by May 31.

With an eigenvector centrality of 1.0, the selected hub node is the most "connected" nursing home in Florida (Fig. 2C). Not only is this facility directly linked to 52 other homes—substantially higher than the state's average of 11.4—many of these direct connections are themselves highly connected, demonstrating the importance of capturing the entire network in these outcome measures. A small number of facilities have disproportionate influence in the overall network in Florida, with only 4% of nursing homes having eigenvector centrality of > 0.5. The hub node in Georgia (Fig. 2D) faced a large COVID outbreak, with 220 confirmed or suspected cases among residents by May 31, one of the largest outbreaks in the state. With a degree

of 34 and relatively high eigenvector centrality of 0.56, this node is directly linked to other highly connected facilities, including several with large COVID outbreaks.

Home to more than 600 skilled nursing facilities, New York State has an average degree of 7.8. While this illustrative hub facility (Fig. 2E) has only five direct connections, these neighbors are highly connected themselves, resulting in a weighted average neighbor degree of 42, well above the state's average. Lastly, a Pennsylvania nursing home (Fig. 2F) has both a degree and strength of 10, meaning that only one smartphone appears in both the focal home and each connected facility. This particular nursing home illustrates how direct connections act as bridges to other clusters of homes, potentially importing or exporting SARS-CoV-2 infection across different subnetworks.

Cross-Sectional Results

Table 3 presents multivariate regressions of cumulative nursing home COVID-19 cases as of May 31 on a set of explanatory variables. Importantly, these regression specifications include state fixed effects to allow for differences in baseline risks and reporting practices across states; we include even finer county fixed effects in *SI Appendix*. We use the inverse hyperbolic sine of cases as the dependent variable, given its nonnegative skewed distribution. Column 1 shows our base specification with our simplest network explanatory variable, *node degree* k_i—the number of "neighbors" or other nursing homes connected to the focal home by at least one smartphone. Results indicate that, if a home adds one neighbor (average degree is 7.1), the expected number of COVID-19 cases increases by $3.43\% \times 4.43 = 15.2\%$. Column 2 replaces the degree measure with *node strength* s_i—the total number of "contacts" or smartphones that appear in the facility of interest and in some other nursing homes. This too predicts nursing home cases significantly: if a home adds one contact (average strength is 8.8), expected cases of COVID-19 increase by $1.63\% \times 2.14 = 3.48\%$. Column 3 replaces the degree and strength measures with *weighted average neighbor degree* $\bar{k}_i^w$—the average degree of a nursing home's neighbors, weighted by strength to the focal node (33). Here, an increase of one (mean is 10.2) leads to an expected $4.10\% \times 2.05 = 8.41\%$ increase in cases.
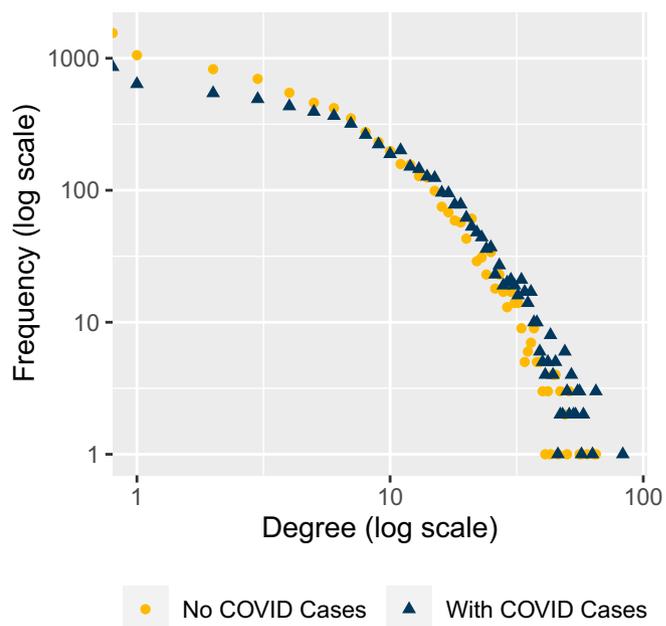

Fig. 1. Degree distribution of nursing homes with and without COVID cases (reported to CMS as of May 31, 2020).
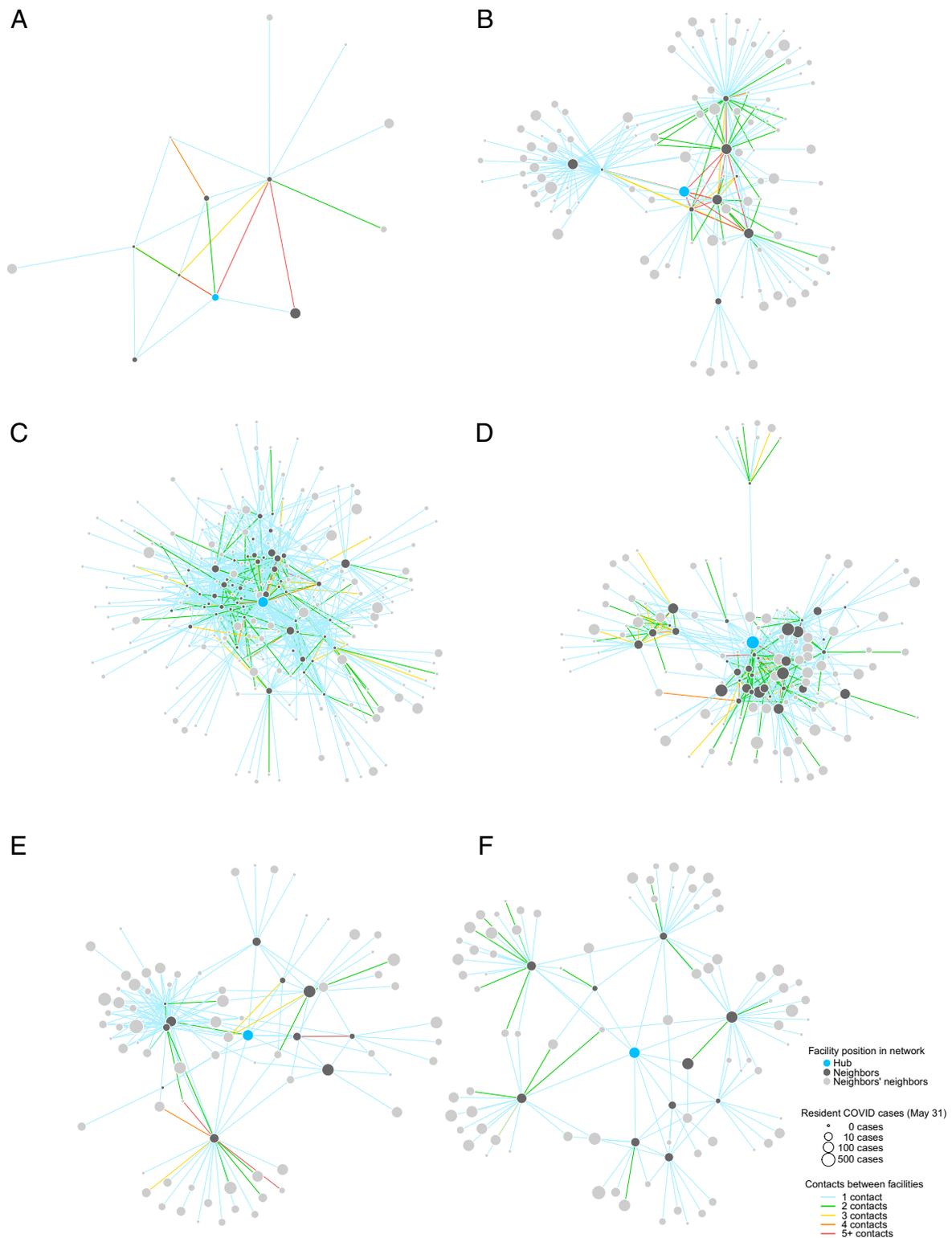
Chen et al.
Nursing home staff networks and COVID-19

PNAS | 3 of 7
https://doi.org/10.1073/pnas.2015455118

Fig. 2. Network structure of selected nursing home facilities in Alabama (*A*), California (*B*), Florida (*C*), Georgia (*D*), New York (*E*), and Pennsylvania (*F*). Details for each hub facility are provided in Table 2.

Column 4 of Table 3 may be of particular interest to policy makers, as it examines the predictive power of local network features, potentially knowable by individual nursing home administrators. Intuitively, this regression compares demographically and geographically situated nursing homes of similar quality, which are thus likely exposed to similar risks of community

spread. Regression 4 suggests that 49% of nursing home resident cases are attributable to shared staff transmitting the virus across multiple nursing homes.

Column 5 of Table 3 uses our final network measure, *eigenvector centrality* v_i—the extent to which a nursing home is connected to other highly connected nursing homes, normalized to range

Table 2. Network centrality measures for six selected nursing homes

Hub facility	State	COVID cases	Degree	Strength	WAND	Eigenvector centrality
A	AL	8	6	56	8.8	<0.01
B	CA	63	9	83	24.1	0.09
C	FL	54	52	81	23.9	1.00
D	GA	220	34	57	24.4	0.56
E	NY	62	5	5	42.4	0.12
F	PA	78	10	10	13.5	0.08

COVID cases are confirmed and suspected cases among residents reported to CMS as of May 31, 2020. WAND, weighted average neighbor degree.

between 0 and 1 within each state. This measure implies that, as we move from an unconnected nursing home in the state ($v_i = 0$) to the most connected ($v_i = 1$), expected cases increase by 195%.

Consistent with other studies (13–15), we find that CMS ratings of nursing home quality are not predictive of infections, yet facilities in urban locations, those with more beds, a higher share of Black residents, or a higher share of residents on Medicaid are all associated with more COVID-19 cases (details in *SI Appendix*, Table S2). We find that for-profit homes are associated with more COVID-19 cases, consistent with ref. 20.

Time Series Evidence

One potential limitation of our analysis is that we do not explicitly show that SARS-CoV-2 travels from home to home. Given data limitations—particularly the late initiation of CMS and state reporting and the failure of many states to archive early reporting—we do not have consistent time series data for individual nursing homes to examine cases over a long time period. However, we were able to hand-collect weekly home-by-home data on the presence of cases starting in mid-April for three states: Florida, Colorado, and Connecticut. The three states have had very different time patterns of COVID cases (*SI Appendix*, Fig. S1).

To investigate whether an initial COVID outbreak in a nursing home is systematically preceded by outbreaks in homes connected to it, we track each nursing home in these states weekly from the week ending April 19, 2020 until the home's first reported COVID case or August 23, 2020, whichever occurs first. Table 4 presents results of a linear probability model with nursing home–week observations. The dependent variable "first outbreak" is coded as 0 before a nursing home's first case and 1 in the week of its first case. The independent variable in the first column is the number of homes connected to the nursing home that had a first case 2 wk before the examined week. Importantly, county times week fixed effects are included. Column 1 demonstrates that a home with more connections to homes with new outbreaks 2 wk prior are more likely to have a first outbreak in a given week, relative to other homes in the county that same week. Connections in the previous week are somewhat less predictive, but still statistically significantly different from zero. Connections to homes experiencing their first outbreak contemporaneously are not predictive at all (the coefficient is negative but statistically insignificant). One might expect that a spurious correlation—the possibility that connected nursing homes are alike in unobserved ways—would most likely manifest in the data with a finding of contemporaneous outbreaks, but we find no evidence of this.[†] The coefficient magnitude implies that a shared contact with a

nursing home experiencing its first COVID case in week $t-2$ is associated with a 2.5 percentage point higher probability of a home reporting a first case in week t. This is large relative to the baseline probability of a first case in any given week, as the mean of the first week indicator in the regression specification is 0.07.

Discussion

Using a large-scale analysis of smartphone location data, we document substantial connections among nursing homes after nationwide visitor restrictions were enacted in March 2020. Consistent with the CDC's conclusion that shared workers were a source of infection for the nursing home outbreak in Kirkland, WA (8), our network measures suggest that staff linkages between nursing homes are a significant predictor of SARS-CoV-2 infections. Our general findings are robust to alternative specifications or the use of the case count data available from CMS. Although one cannot conclusively draw causal inferences from an observational study, this is an environment in which randomized controlled trials, natural experiments, or other causal methodologies are not readily available.

These results provide evidence for a policy recommendation of compensating nursing home workers to work at only one home and limit cross-traffic across homes. While some nursing homes and other long-term care facilities have undertaken actions to create a "staff bubble," this is still not a component of extant regulation (34, 35). Absent such regulation, allocation of PPE, testing, and other preventive measures should be targeted thoughtfully, recognizing the current potential for transmission across homes. New CMS testing guidelines as of August 2020 state that a nursing home not experiencing a current outbreak and located in a county with case positivity rates of less than 5% need only test staff members once per month (12). If two homes are known to share workers, however, testing could be increased at one home if an outbreak occurs at the other facility. Further, given the greater chance that a highly connected home experiences a new outbreak—and the risk this creates for its connections—more frequent testing of highly connected homes could be warranted, even when county positivity rates are low. While the nursing home population is particularly fragile, this research has implications for cross-linkages in other congregate settings such as assisted living homes, prisons, food-processing plants, and large workplace facilities.

Materials and Methods

Smartphone Location Data. We estimate staff and contractor networks across nursing homes using anonymized smartphone location data provided by Veraset, a company that aggregates location data across several apps on both the Apple and Android platforms after the user consents to the use of their anonymized data. While these methods contain geolocation information from only a subset of all smartphones, previous studies with these data have found them to be highly representative of the United States on numerous demographic dimensions (36). A smartphone typically reports ("pings") a user's location every 10 min throughout the day. We filter these data to estimate user/nursing home visits by, first, excluding visits with fewer than three user pings inside that home that day, then further excluding visits whose first and last pings are separated by less than 1 h. This helps reduce staff false-positives due to GPS error or users who briefly enter a home (like a delivery person). Under this definition, of the more than 50 million smartphones in our US sample, we identify 501,503 smartphones that visit at least one US nursing home between March 13 and May 28, 2020, and a visitor to a home visits that home an average of 16 d over our 11-wk study period.

We match all US nursing homes with a shapefile delineating each facility's rooftop boundary. To do so, we match a nursing home's CMS-provided street address to a latitude–longitude location using the Google Maps Application Programming Interface (API), and then match that location to a satellite image machine-learned geofence of the convex hull of the building's rooftop (provided by Microsoft/ Open Streetmaps). Using these rooftop geofences, we find all times that a sampled smartphone spends more than 1 h in a US nursing home during our

---

[†]The exact time lag anticipated by a shared staff mechanism is unclear. Given the weekly reporting cadence, the typical 2-wk infectivity period, and the fact that a shared staff member who carries the disease from home A to home B may not be the index case at home A, we expect some lag from measured and reported cases at home A to measured and reported cases at home B.

Chen et al.
Nursing home staff networks and COVID-19

PNAS | 5 of 7
https://doi.org/10.1073/pnas.2015455118

Table 3. Covariates of COVID-19 cases within nursing homes

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: $sinh^{-1}$(cases)				
Node degree	0.0343***			0.0242***	
	(0.00255)			(0.00508)	
Node strength		0.0163***		−0.00610+	
		(0.00166)		(0.00297)	
Weighted average neighbor degree			0.0409***	0.0299***	
			(0.00267)	(0.00344)	
Eigenvector centrality in state					1.044***
					(0.109)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,337	6,337	6,337	6,337	6,337
F statistic	123.4	114.9	128.7	112.9	114.5
R^2	0.408	0.400	0.412	0.415	0.399
Within R^2	0.189	0.178	0.195	0.199	0.177

Standard errors are in parentheses. Significance levels: $^+p < 0.05$, $^*p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$. Dependent variable is inverse hyperbolic sine of COVID cases using state data. Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid. CMS quality is a 1 to 5 categorical rating.

study period, when visitor restrictions were in effect. By identifying smartphones that entered more than one nursing home, we measure the nursing home staff contact network.

Network Metrics. The contact structure among nursing homes within a state is represented by an undirected network consisting of n nodes (the facilities) and $n(n-1)/2$ possible edges (pairs of facilities). We construct a symmetric $n \times n$ adjacency matrix $\mathbf{A}$, where $a_{ij} = 1$ if at least one smartphone is observed in both facilities i and j, and 0 otherwise. Edge weights w_{ij} correspond to the number of smartphones observed in both facilities. State-level summaries of each network measure are in *SI Appendix*, Table S1.

A facility's degree k_i equals the total number of other nursing homes connected to facility i (i.e., the number of node i's neighbors).

$$k_i = \sum_{j=1}^{n} a_{ij}. \quad \text{[1]}$$

Strength s_i is the weighted sum of contacts with other facilities (i.e., the total number of smartphones that appear in facility i and some other nursing home).

$$s_i = \sum_{j=1}^{n} w_{ij} a_{ij}. \quad \text{[2]}$$

Weighted average neighbor degree $\bar{k}_i^w$ is the average degree of node i's neighbors (i.e., the neighbors' connections to other facilities), weighted by the number of connections w_{ij} shared with node i, as previously defined (33).

$$\bar{k}_i^w = \frac{1}{s_i} \sum_{j=1}^{n} w_{ij} a_{ij} k_j. \quad \text{[3]}$$

Eigenvector centrality v_i measures the extent to which node i is connected to other highly connected nodes in the network.

$$v_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} v_j. \quad \text{[4]}$$

This measure is computed using the principal eigenvector of the adjacency matrix, rewritten in matrix notation as $\mathbf{Av} = \lambda\mathbf{v}$. We normalize v_i to range between 0 and 1 within each state.

Cross-Sectional Regressions. Our main specification examines predictors of nursing home resident COVID cases as a function of several explanatory variables. We include the home's demographic characteristics, including linear and quadratic terms for the number of beds. Following previous literature (13), we include indicator variables for whether a nursing home has

a large proportion ($> 50\%$) of residents on Medicaid and a large proportion ($> 25\%$) of Black residents. We include CMS quality measures, as done previously (13), which rate nursing homes on a five-point scale; we include this as a categorical variable, with the omitted category being five stars (the highest possible rating). We include an indicator variable if the home had infection control violations in its most recent inspection. Finally, we define an indicator variable for whether a home is in an urban location based on the CDC's urban–rural classification (37).

To examine whether nursing home connectivity predicts COVID-19 cases, we use the following regression model:

$$sinh^{-1}(Cases_i) = \beta_0 + \beta_1 NodeDegree_i + \beta_2 NodeStrength_i$$
$$+ \beta_3 WeightNeighDeg_i + \beta_4 EigenCentrality_i$$
$$+ \gamma_0 \mathbf{X}_i + \gamma_1 \mathbf{F}_i + \varepsilon_i,$$

where $sinh^{-1}(x) = \ln(x + \sqrt{1+x^2})$ is the inverse hyperbolic sine of a nursing home's COVID cases. All reported semielasticities are adjusted for the $sinh^{-1}$ functional form. We include, as independent variables, the four network measures that characterize a home's connectivity, as described in the previous section. The vector $\mathbf{X}_i$ includes demographic, geographic, and regulatory controls for nursing home i. To control for reporting and other differences across states, we include state fixed effects denoted by $\mathbf{F}_i$. Coefficients for all variables are in *SI Appendix*, Table S2, with results for individual states

Table 4. Time series evidence

	(1)	(2)	(3)
	Dependent variable: First outbreak indicator		
New outbreak degree$_{t-2}$	0.0245*		
	(0.00810)		
New outbreak degree$_{t-1}$		0.0154+	
		(0.00680)	
New outbreak degree$_t$			−0.0128
			(0.00972)
Fixed effects	County × week	County × week	County × week
Observations	7,429	7,429	7,429
F statistic	9.142	5.156	1.7415
R^2	0.213	0.212	0.211
Within R^2	0.00309	0.00138	0.000412

Standard errors are in parentheses. Significance levels: $^+p < 0.05$, $^*p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$. Dependent variable is a binary variable that equals 1 for the nursing home-week in which a home first reports having a COVID case using state data for Colorado, Connecticut, and Florida.

reported in *SI Appendix*, Table S3. *SI Appendix*, Table S4 replaces the dependent variable with a binary indicator variable if the nursing home has had any COVID cases. *SI Appendix*, Table S5 repeats our analysis using county fixed effects, an important robustness check given that county-level SARS-CoV-2 prevalence predicts case counts in nursing facilities (15). Finally, *SI Appendix*, Table S6 replaces the data for 22 states with the larger CMS dataset for the continental United States.

Time Series. For the time series analysis, we use daily nursing home resident case counts provided by the Florida, Connecticut, and Colorado Departments of Health (38–40). Because the states did not all report cumulative cases, we construct a binary variable first outbreak to indicate the first week in which each nursing home appears in the database. The case data for all three states begin by the week ending April 19; thus, we construct network measures from 5 wk of smartphone data, beginning with the visitor lockdown after March 13 until April 19, 2020.

We use a linear probability model, with each observation a nursing home-week. For each home i in week t, we regress the "first outbreak" indicator on county-week interaction fixed effects and on $NewOutbreakDegree_{i,t-2}$, the number of nursing homes directly connected to home i that experienced their first recorded case in week $t-2$. For example, if home A is only connected to home B, and home B experiences its first outbreak during the week of May 3, 2020, the variable $NewOutbreakDegree_{A,t-2}$ equals zero in all weeks except the week of May 17, 2020. $NewOutbreakDegree_{i,t-2}$ is similar to our previous measure $NodeDegree_i$, but is calculated only for

homes with reported outbreaks and is lagged by 2 wk. With the inclusion of county-week fixed effects, the specification is effectively estimating whether a home i in county j in week t is differentially likely to have an outbreak in week t relative to other homes in county j and week t if i is connected to homes that had first outbreaks in week $t-2$. We obtain the following specification:

$$FirstOutbreak_{i,t} = \beta_0 + \beta_1 NewOutbreakDegree_{i,t-2}$$
$$+ \gamma_1 \mathbf{F}_{i,t} + \varepsilon_{i,t}.$$

We repeat this specification using connections to homes in the previous week using $NewOutbreakDegree_{i,t-1}$, and contemporaneous connections using $NewOutbreakDegree_{i,t}$. We exclude from our analysis any homes that have connections to homes out of state, as we cannot observe detailed case timing for homes in states neighboring these three states.

Data Availability Statement. Nursing home network data and replication code will be available on the website of author M. Keith Chen, as well as archived on Harvard Dataverse at https://doi.org/10.7910/DVN/FTWI83.

1. M. Conlen *et al.*, More than 40% of U.S. coronavirus deaths are linked to nursing homes. *New York Times.* https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html. Accessed 15 July 2020.
2. Kaiser Family Foundation, State data and policy actions to address coronavirus, https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus. Accessed 11 July 2020.
3. G. Girvan, A. Roy, Foundation on Research on Equal Opportunity: Long term care facilities reported COVID-19 deaths by state. https://docs.google.com/spreadsheets/d/17JmyFzOd3ZEYCGpP0mK5lS_P10yPgCuYF8PYALKuTs8/edit#gid=0. Accessed 31 August 2020.
4. L. J. Strausbaugh, S. R. Sukumar, C. L. Joseph, K. P. High, Infectious disease outbreaks in nursing homes: An unappreciated hazard for frail elderly persons. *Clin. Infect. Dis.* **36**, 870–876 (2003).
5. L. E. Lansbury, C. S. Brown, J. S. Nguyen-Van-Tam, Influenza in long-term care facilities. *Influenza Other Respir. Viruses* **11**, 356–366 (2017).
6. Centers for Medicare and Medicaid Services, COVID-19 nursing home data as of week ending 6/28/2020. https://data.cms.gov/stories/s/COVID-19-Nursing-Home-Data/bkwz-xpvg/. Accessed 12 July 2020.
7. H. D'Adamo, T. Yoshikawa, J. Ouslander, Coronavirus disease 2019 in geriatrics and long-term care: The ABCDs of COVID-19: COVID-19 in geriatrics and long-term care. *J. Am. Geriatr. Soc.* **68**, 912–917 (2020).
8. T. M. McMichael, COVID-19 in a long-term care facility—King County, Washington, February 27–March 9, 2020. *Morb. Mortal. Weekly Rep.* **69**, 339–342 (2020).
9. Centers for Medicare and Medicaid Services, Guidance for infection control and prevention of coronavirus disease 2019 (COVID-19) in nursing homes. https://www.cms.gov/files/document/3-13-2020-nursing-home-guidance-covid-19.pdf. Accessed 12 July 2020.
10. T. M. McMichael *et al.*, Epidemiology of Covid-19 in a long-term care facility in King County, Washington. *N. Engl. J. Med.* **382**, 2005–2011 (2020).
11. Centers for Medicare and Medicaid Services, Guidance for infection control and prevention of coronavirus disease 2019 (COVID-19) in nursing homes. https://www.cms.gov/files/document/4220-covid-19-long-term-care-facility-guidance.pdf. Accessed 10 July 2020.
12. Centers for Medicare and Medicaid Services. Interim final rule (IFC), CMS-3401-IFC, additional policy and regulatory revisions in response to the COVID-19 public health emergency related to long-term care (LTC) facility testing requirements and revised COVID19 focused survey tool. https://www.cms.gov/files/document/qso-20-38-nh.pdf. Accessed 26 August 2020.
13. H. R. Abrams, L. Loomer, A. Gandhi, D. C. Grabowski, Characteristics of US nursing homes with COVID-19 cases. *J. Am. Geriatr. Soc.* **68**, 1653–1656 (2020).
14. R. T. Konetzka, Senate testimony before the senate special committee on aging. https://www.aging.senate.gov/imo/media/doc/SCA_Konetzka_05_21_20.pdf. Accessed 10 July 2020.
15. E. M. White *et al.*, Variation in SARS-CoV-2 prevalence in US skilled nursing facilities. *J. Am. Geriatr. Soc.* **68**, 2167–2173 (2020).
16. K. Pillemer, L. Subramanian, N. Hupert, The importance of long-term care populations in models of COVID-19. *J. Am. Med. Assoc.* **324**, 25–26 (2020).
17. N. Ferguson *et al.*, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand" (Rep. 9, Imperial College London, 2020).
18. M. He, Y. Li, F. Fang, Is there a link between nursing home reported quality and COVID-19 cases? Evidence from California skilled nursing facilities. *J. Am. Med. Dir. Assoc.* **21**, 905–908 (2020).
19. Y. Li, H. Temkin-Greener, S. Gao, X. Cai, COVID-19 infections and deaths among Connecticut nursing home residents: Facility correlates. *J. Am. Geriatr. Soc.* **68**, 1899–1906 (2020).
20. P. Rowan *et al.*, "A study of the Covid-19 outbreak and response in connecticut long-term care facilities" (Tech. Rep. DPH 2021-0041, Mathematica Policy Research, 2020).
21. S. N. Ladhani, J. Y. Chow, R. Janarthanan *et al.*, Increased risk of SARS-CoV-2 infection in staff working across different care homes: Enhanced CoVID-19 outbreak investigations in London care homes. *J. Infect.* **81**, 621–624 (2020).
22. F. A. Tobolowsky *et al.*, Covid-19 outbreak among three affiliated homeless service sites—King County, Washington, 2020. *Morb. Mortal. Weekly Rep.* **69**, 523–526 (2020).
23. J. W. Dyal *et al.*, COVID-19 among workers in meat and poultry processing facilities–19 states, April 2020. *Morb. Mortal. Weekly Rep.* **69**, 557–561 (2020).
24. S. A. Kinner, *et al.*, Prisons and custodial settings are part of a comprehensive response to COVID-19. *Lancet Public Health* **5**, e188–e189 (2020).
25. T. Williams, R. Griesbach, San Quentin Prison was free of the virus. One decision fueled an outbreak. *New York Times*, 30 June 2020. https://www.nytimes.com/2020/06/30/us/san-quentin-prison-coronavirus.html. Accessed 10 July 2020.
26. L. L. Stanley, Influenza at San Quentin Prison, California. *Publ. Health Rep.* **34**, 996–1008 (1919).
27. New York State Department of Health, Factors associated with nursing home infections and fatalities in New York state during the COVID-19 global health crisis. https://health.ny.gov/press_releases/2020/docs/nh_factors_report.pdf. Accessed 11 July 2020.
28. V. W. Slaugh, A. A. Scheller-Wolf, S. R. Tayur, Consistent staffing for long-term care through on-call pools. *Prod. Oper. Manag.* **27**, 2144–2161 (2018).
29. S. F. Lu, L. X. Lu, Do mandatory overtime laws improve quality? Staffing decisions and operational flexibility of nursing homes. *Manag. Sci.* **63**, 3566–3585 (2017).
30. N. G. Castle, Use of agency staff in nursing homes. *Res. Gerontol. Nurs.* **2**, 192–201 (2009).
31. R. O. Cadigan, D. G. Stevenson, D. J. Caudry, D. C. Grabowski, Private investment purchase and nursing home financial health. *Health Serv. Res.* **50**, 180–196 (2015).
32. C. H. Van Houtven, N. DePasquale, N. B. Coe, Essential long-term care workers commonly hold second jobs and double- or triple-duty caregiving roles. *J. Am. Geriatr. Soc.* **68**, 1657–1660 (2020).
33. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
34. C. Sudo, How 3 senior living communities 'bubbled' communities by having staff live on site. *Senior Housing News*, 11 May 2020. https://seniorhousingnews.com/2020/05/11/how-3-senior-living-providers-bubbled-communities-by-having-staff-live-on-site/. Accessed 18 July 2020.
35. D. Rodricks, *How a west Baltimore nursing home has zero COVID infections*. *Baltimore Sun*, 19 June 2020. https://www.baltimoresun.com/opinion/columnists/dan-rodricks/bs-md-rodricks-0619-20200618-re2obahhbzbddojazrkf65gdoq-story.html. Accessed 18 July 2020.
36. M. K. Chen, R. Rohla, The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).
37. Centers for Disease Control and Prevention, NCHS urban–rural classification scheme for counties. https://www.cdc.gov/nchs/data_access/urban_rural.htm. Accessed 12 July 2020.
38. Florida Department of Health, COVID-19 report archive. ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/. Accessed 26 September 2020.
39. Connecticut Department of Public Health, Nursing homes with residents positive for covid-19. https://data.ct.gov/Health-and-Human-Services/Nursing-Homes-with-Residents-Positive-for-COVID-19/wyn3-qphu. Accessed 26 September 2020
40. Colorado Department of Public Health and Environment, Outbreak data archive. https://covid19.colorado.gov/covid19-outbreak-data. Accessed 26 September 2020.

Chen et al.
Nursing home staff networks and COVID-19

PNAS | 7 of 7
https://doi.org/10.1073/pnas.2015455118