PNAS NEXUS

PNAS Nexus, 2025, **4**, pgaf180
https://doi.org/10.1093/pnasnexus/pgaf180
Advance access publication 9 June 2025
Research Report

Prison connectivity and disease transmission to neighboring communities: The role of prison staff

Yilin Zhuo[a], Kristin Turney[b], Naomi F. Sugie[c,*], Emily Owens[d,e] and M. Keith Chen[a]

[a]Anderson School of Management, University of California Los Angeles, Los Angeles, CA 90095, USA
[b]Department of Sociology, University of California Irvine, Irvine, CA 92697, USA
[c]Department of Sociology, University of California Los Angeles, Los Angeles, CA 90095, USA
[d]Department of Criminology, Law and Society, University of California Irvine, Irvine, CA 92697, USA
[e]Department of Economics, University of California Irvine, Irvine, CA 92697, USA
*To whom correspondence should be addressed: Email: nsugie@soc.ucla.edu
Edited By Eric Schneider

Abstract

Using smartphone location data and a novel application of publicly available employment data, we map how California communities are connected to nearby prisons through the movement of prison staff, and we measure the role these connections play in spreading infectious diseases. Leveraging an exogenous prisoner transfer-induced COVID-19 outbreak at San Quentin state prison in June 2020 as a quasiexperiment, we examine the unidirectional spread of the disease from the prison to surrounding communities. This outbreak was unique: its origin from outside Northern California was clearly documented and nonstaff entry and exit was severely limited during this time. Our identification strategy compares zip codes connected and unconnected to the prison via staff movement. Compared to unconnected zip codes with similar pretransfer COVID-19 rates and demographic characteristics (race/ethnicity, education, household income, age, and population), zip codes connected to San Quentin had 13% more new COVID-19 cases in July and 30% more in August. Our results suggest that a hypothetical novel infectious disease that emerged in California prisons could lead to almost 15,000 community infections within 1 month from staff movements alone. These findings identify the degree to which "closed institutions" are—even during lockdowns—epidemiologically porous, highlighting the need for public health interventions to reduce the unintended consequences of such connections on the spread of infectious disease.

Keywords: prisons, incarceration, staff networks, infectious diseases, community risk

Significance Statement

It is well recognized that "closed institutions" like prisons pose infectious disease health risks to their incarcerated residents, especially during outbreaks and epidemics. Less well understood is to what degree vulnerabilities within prisons spread to surrounding communities and heighten overall epidemic risks. Using smartphone data to identify and track staff movements during a prisoner transfer-induced COVID-19 outbreak at San Quentin prison, we find evidence that even during lockdowns, staff transmit COVID-19 cases to their communities, and we show that similar patterns can be found using readily available data from the Census Bureau. These findings suggest the importance of targeted public health interventions in prisons, as infectious diseases can quickly spread not only within prisons but to their broader local population.

Introduction

In Brown v. Plata 131 S. Ct. 1910, 563 US 493 (2011), the US Supreme Court confirmed what many scholars and advocates had long argued—that the American criminal legal system creates a substantial public health burden. People who have contact with the criminal legal system—through arrest, conviction, and incarceration—have worse health outcomes along almost every dimension (1), with consequences for their community's health. Yet, prison staff—and their frequent movement between prisons and communities—is an often under-recognized, but potentially important, additional vector of disease transmission (2–4). Two recent revolutions in technology and global health—the rise of geographic information system (GIS)-enabled smartphones and the COVID-19 pandemic—allow us to directly identify how prison staff are connected to surrounding communities and the degree to which these connections transmit upper respiratory diseases to communities.

The largest COVID-19 outbreaks occurred in "closed institutions," such as prisons, jails, and nursing homes (5, 6). These facilities concentrate various vulnerabilities—e.g. confined physical

Competing Interest: The authors declare no competing interests.
Received: October 7, 2024. **Accepted:** March 31, 2025

spaces and overcrowding, limited resources and personal protective equipment, and individuals with socioeconomic disadvantages and health comorbidities—into places of extreme risk. Indeed, US prisons were the sites of 39 of the country's 50 largest COVID-19 outbreaks in 2020 (6). People incarcerated in prisons, compared to the general population, were five times as likely to experience COVID-19 infection and three times as likely to die from COVID-19 (7).

The extent of disease transmission from prison COVID-19 outbreaks to the general population is largely unknown. On the one hand, prisons may pose little risk to the general population. Beginning in March 2020, prisons across the country locked down their facilities, prohibiting visitors and programming. Although state agencies made different decisions regarding when to reopen (8), they appeared to have responded to general population case rates, closing back down when infections were rising in facilities and communities. On the other hand, despite these lockdowns, prisons are more porous than many realize, with staff continually moving back and forth between their homes and workplaces (4). Notably, in most jurisdictions, agencies did not implement universal staff testing for COVID-19 (9), and mask mandates were not consistently enforced (10); consequently, prison staff experienced higher positivity rates compared to the general population (11, 12).

The role of prison staff movements in COVID-19 disease transmission to their surrounding communities is challenging to identify and isolate from other factors. Studies have shown that community rates of COVID-19 are related to nearby prison and jail COVID-19 rates (2, 3, 13, 14), that other infectious diseases move in and out of prisons (15), and that correctional staff are likely important vectors of infection in simulation models (16). However, observational studies are commonly limited by endogeneity issues, and correlations between community and prison infections do not necessarily identify prison staff as transmission vectors, as opposed to other transmission mechanisms, such as recent releases of incarcerated people to their communities (9, 10, 17) and routine contacts between incarcerated people and communities (e.g. for medical care; 18). Moreover, because prison staff likely serve as vectors of both disease entry and exit between communities and prisons, isolating the causal pathway of disease transmission from prisons to communities (as opposed to the reverse relationship of communities to prisons) is not possible in most observational studies.

We leverage multiple sources of "big data," specifically smartphone location data purchased from Veraset and publicly available prison property boundary line data from the Department of Homeland Security (DHS) to provide evidence on how connected different California communities are to prisons and how those connections facilitate disease transmission from prisons to communities. These data include high-quality and fine-grained information on which individuals travel to and from prisons and how long they spend in prisons and nonprison locations. We also replicate these analyses with publicly available LODES (LEHD Origin–Destination Employment Statistics) data from the Census Bureau, which contain information on the communities where workers covered by unemployment insurance live, measured annually and over all physical work locations, to illustrate how researchers can effectively model prison connectivity and disease transmission in neighboring communities in future research using either data source.

The smartphone GPS data have been previously used to document the mobility of nursing home employees (19), police officers (20), and tech firm employees (21). The data include "pings" indicating the location of a smartphone at a particular point in time and capture 10 to 20% of the US population. Pings are logged whenever a participating smartphone application requests location information, with the modal time between a phone's two consecutive pings being roughly 10 min. To identify people who work in prisons, we first define a set of small geographic regions (~153 m × 153 m grid cells called "geohash-7" areas) that overlap with or are adjacent to prison boundaries. We then identify all smartphones that ping within one of those small areas for at least 10 min within a single 30-min period. This filters out any transient phones that, for example, are driving on a nearby road, and ensures that we identify people who genuinely stop at the prison. We then identify and measure connectivity between the prison and zip codes by associating the total time spent in prisons for all smartphones that "live" in a certain zip code. We define the "home" of a smartphone as the zip code that the phone spends the most time when outside of prisons (in June to October 2020). Appendix Table S1 presents the number of identified phones for 35 state prisons in June 2020, ranging from 60 (California Correctional Center) to 648 (California Institution for Men). Most prisons had between 100 and 200 identified phones. Additionally, we compare these smartphone-derived staff counts to LODES-based counts; Fig. 1 shows a strong and positive correlation, supporting the validity of our smartphone-based measure of prison staff.

It is likely that prison staff, rather than visitors to prisons, drive these community connections for three reasons. First, as a COVID-19 mitigation policy, California prisons prohibited in-person visitors during this time. Second, zip codes connected to a specific prison are generally close to that specific prison; for example, zip codes connected to San Quentin are generally neighboring areas close to San Quentin (Fig. 2B); prison staff are more likely to live close to their place of employment than visitors are to live close to their incarcerated family member or friend. Third, the zip codes that our smartphone-based measurement identifies as connected to prisons do not closely correspond with the zip codes where most prisoners lived prior to incarceration (22). For example, Fig. 2A reveals that the greater Los Angeles area is only loosely connected to prisons on a day-to-day basis during the early pandemic even though it is the most heavily populated region of California and is the largest contributor to the California prison population. Indeed, the correlation between zip codes of prisoner origin and zip codes of staff connections is low (0.191). Importantly, these smartphone data show, for the first time, which California zip codes were the most connected to prisons in June 2020 (Fig. 2).

Quantifying how prisons impact community health requires differentiating transmission of infection from communities into prisons from transmission from prison into communities. In this article, we leverage the fact that the source of one prominent COVID-19 outbreak in San Quentin State Prison, in Northern California, was exogenous to that facility and local COVID-19 conditions. This outbreak was caused by transferring 122 incarcerated people from the California Institution for Men, in Southern California, to San Quentin while the former facility was experiencing an outbreak (23). Transferred prisoners tested negative for COVID-19 more than 2 weeks before being transferred, were placed on overcrowded buses for their transport, and 15 of those transferred tested positive for COVID-19 shortly after arriving at San Quentin. Prison administrators housed the remaining transfers in a housing unit without solid doors and allowed staff to work in multiple areas throughout the prison, all of which contributed to a massive outbreak among prisoners at that facility (Fig. 3).

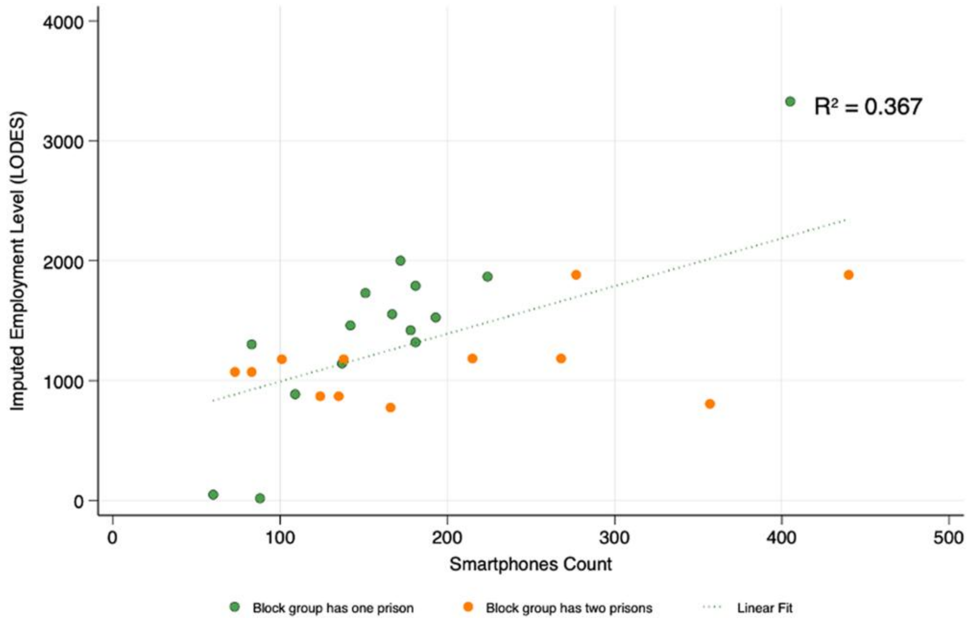

Fig. 1. Correlation between prison-level LODES staff counts and smartphone-derived staff measures. The scatterplot displays the relationship between aggregated smartphone counts of prison staff from 27 state prisons and the imputed employment level derived from LODES (LEHD-Original Destination Statistics) data. The LODES staff count is estimated based on the total number of jobs in the block groups where prisons are located. If a block group contains two prisons, the imputed employment level for each prison is calculated by dividing the total block group job count by the number of prisons. LODES data is unavailable for the block groups where eight prisons are located, leading to missing estimates for these facilities.

Alongside the transfer to San Quentin, it is important to note that Corcoran State Prison received 67 incarcerated people from California Institution for Men and experienced a more limited outbreak among its incarcerated population. Corcoran is a newer prison compared to San Quentin, with many cells having solid doors believed to limit the spread of respiratory infection among incarcerated people (23). We focus on the larger San Quentin outbreak, but we also report results on community transmission from Corcoran in the Materials and methods section and Appendix; while the internal Corcoran outbreak was smaller, and Corcoran is not identified in the LODES data, we find slightly larger rates of transmission into the community via prison staff.

The California Office of the Inspector General concluded that these policy and implementation failures caused a "public health disaster at San Quentin State Prison" (10); in this study, we identify how these failures also endangered the public health of surrounding communities, extending beyond the prison walls of San Quentin. Using monthly counts of zip code-level new COVID-19 cases from the California Department of Public Health (CDPH), we quantify the importance of prison connectivity in disease transmission to neighboring communities by testing whether this specific prison outbreak at San Quentin led to a differential increase in COVID-19 risk in the specific zip codes most connected to San Quentin prison due to staff movement.

Our empirical strategy compares the monthly COVID-19 case rates following the transfer between connected zip codes (treated) and unconnected zip codes (control) that have similar pretransfer zip code-level COVID-19 case rates (obtained from the CDPH) and demographic characteristics (race/ethnicity, education, household income, age, and population from the 2015–2019 American Community Survey [ACS]), using an unconfoundedness-type of strategy (a "doubly robust" [DR] matching method, see Materials and methods section) to account for nonlinear COVID-19 transmission (24). We define "connected" zip codes as areas where our smartphone data suggest a connection to San Quentin. There are 93 zip codes that are connected to San Quentin, all relatively close to the prison. Appendix Table S2 suggests that, on average, these zip codes are more dense, have fewer non-Hispanic white residents (46%), have higher household incomes, and have higher educational attainment than the unconnected zip codes in California. Additionally, they had fewer COVID-19 cases in May 2020 compared to unconnected zip codes. Accordingly, we match and weight treated and control zip codes on pretransfer COVID-19 cases and demographic variables.

Results

The data show that, 1 month after prisoners were transferred to San Quentin, connected zip codes experienced 62 (SE = 23.9) additional COVID-19 cases per 100,000 people, and 2 months later this increased to 81 (SE = 20.6) additional cases, before the rate of new cases converged to the rate in unconnected zip codes. To put these numbers in perspective, when added to the actual rate of new COVID-19 infections in unconnected zip codes, our estimated average treatment effects show that connected zip codes had 13% more new cases in July 2020 and 30% more new cases in August 2020 (Fig. 4 and Appendix Table S3). This is roughly equivalent to the 15% increase in COVID-19 cases observed over 11 weeks when a nursing home was "connected" to a new facility by a shared employee in mid-2020 (19), and is similar to the 13% increase in COVID-19 cases attributed to people cycling in and out of Chicago's Cook County jail in early 2020 (13). Replicating these analyses using publicly available LODES data describing communities where paid prison staff live provides similar results, corresponding to 23% more new cases in connected zip codes in July 2020 and 37% more new cases in August 2020 (Fig. 4B, also see Appendix Table S3). Both the LODES and our smartphone data can be thought of as containing measurement error in the precise movements of all staff, leading our estimates of COVID-19 transmission to be lower bounds of the impact of that movement on disease transmission.

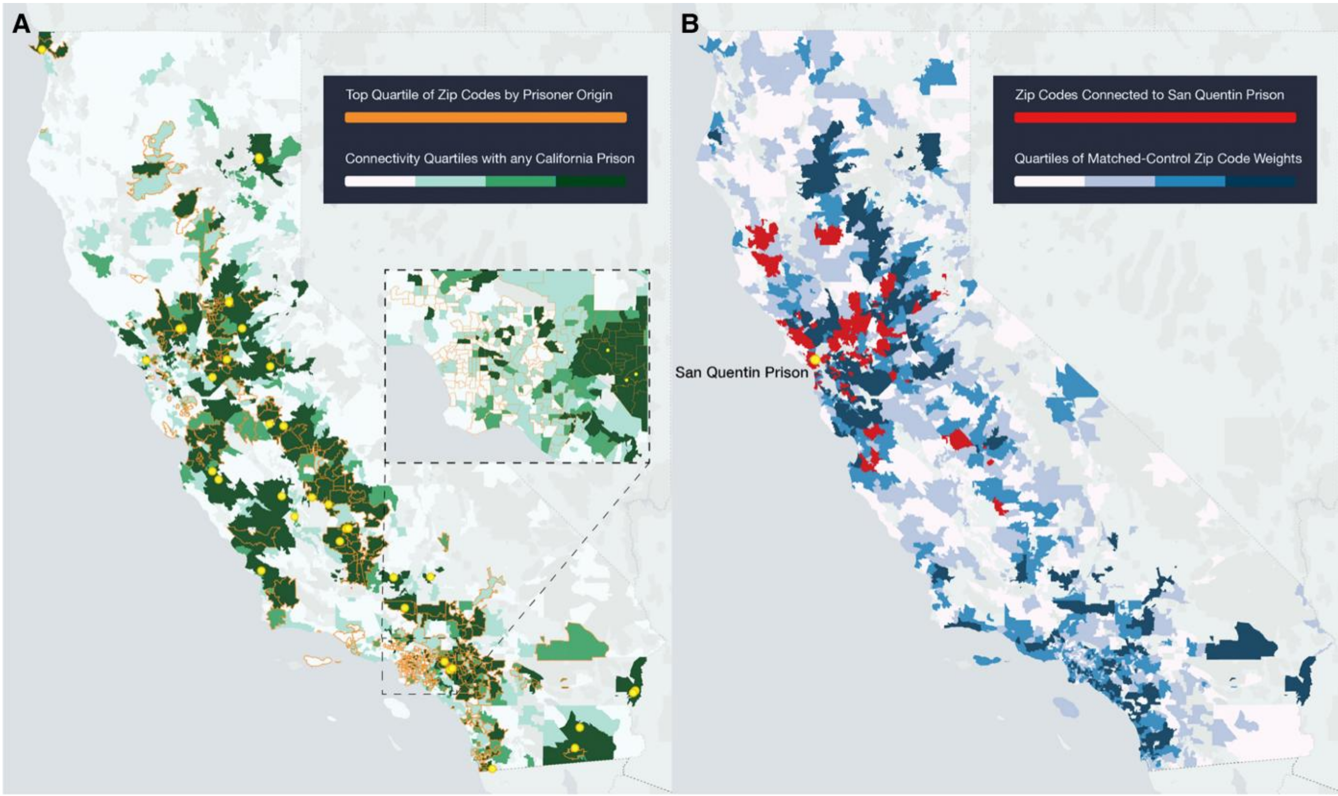

Fig. 2. Zip code connectivity to California state prisons in June 2020. A) Overall connectivity (quartiles in green) to all 35 California state prisons (shown with yellow dots) in June 2020 and the top quartile of prisoner origin (outlined in orange), with an inset showing the Los Angeles region is displayed. B) Zip codes connected (red) or not connected (blue) to San Quentin State Prison in June 2020, with darker blue indicating higher regression-weighting for unconnected zip codes as matched-controls in our analysis are highlighted. All connected prisons have the same weight equal to one.

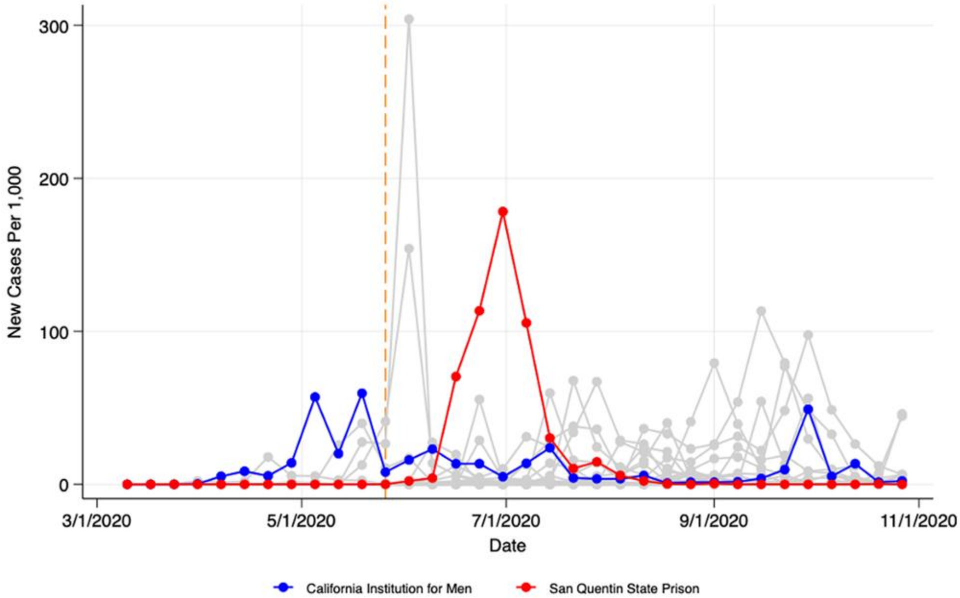

Fig. 3. Weekly COVID-19 cases per 1,000 in California state prisons. The blue line plots new COVID-19 cases per 1,000 people for California Institution for Men, the red line plots new COVID-19 cases per 1,000 people for San Quentin State Prison, and the gray lines plot new COVID-19 cases per 1,000 people for other state prisons. The orange line indicates the date (2020 May 28) when the prison transfer happened.

Of course, even when using a DR matching method, this increased rate of disease transmission could reflect the progression of COVID-19 in zip codes that were connected to prisons generally, rather than San Quentin specifically. This could be the case if zip codes that were demographically similar to zip codes connected to San Quentin, but not connected to any prison, also had some other characteristic not reflected in our census data covariates that drove COVID-19 infections in mid-2020. Examples of this could

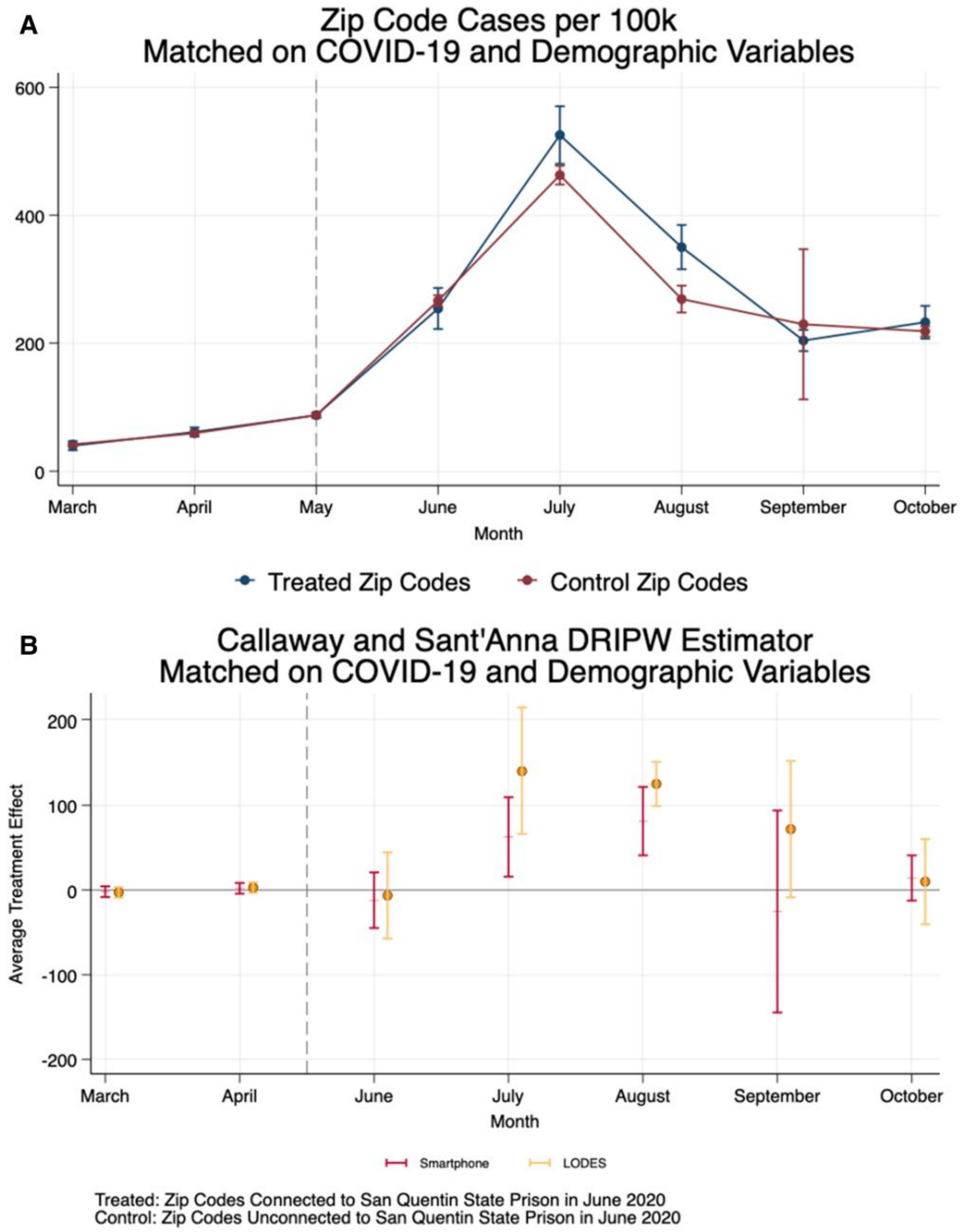

Fig. 4. Monthly COVID-19 case rate and ATE of connecting to San Quentin State Prison. A) The average monthly COVID-19 cases per 100,000 people for connected and unconnected zip codes defined using the smartphone location data, with COVID-19 case rates weighted by the matching-adjustment weights for unconnected zip codes are shown. B) The ATE of connecting to San Quentin State Prison in each month, both estimated using the smartphone location data and using the LODES data is presented. The gray dashed line represents baseline period (May 2020).

include the ability to comply with shelter-in-place orders (25), propensity to adopt personal protective equipment (26, 27), or political affiliation (28).

We explore the potential role of unobserved confounding variables by permuting our measure of connectivity across zip codes and reestimating our average treatment effect (ATE). The distribution of possible effect sizes suggests that the observed timing of COVID-19 transmission is unique to communities connected to San Quentin, rather than communities connected to any other California prison. A total of 34 out of 1,000 estimates for August generated from this permutation test were larger than 81 (Fig. 5).

Additional robustness tests—where we match only on pre-transfer COVID-19 case rates (and not demographics), exclude outlier zip codes, vary how we define the treated zip codes, include interactions between COVID-19 and demographic variables for matching, and examine zip codes connected to Corcoran prison (which experienced a more limited outbreak resulting from a transfer from California Institution for Men)—are included in the Appendix. None of these alternate specifications substantively change our findings, suggesting that the data, rather than any modeling assumptions and/or the uniqueness of the San Quentin case, are driving our estimates.

The fact that we identify a similar—and larger—relationship in Corcoran prison requires investigation, especially because infrastructure-based COVID-19 mitigation strategies (specifically, the ability to house transferred prisoners in cells with solid doors) were considered at the time to be more effective in Corcoran than in San Quentin (10). We believe there are three possible explanations for larger effect sizes in the Corcoran analyses: (i) differences in preexisting COVID-19 case rates in places connected to

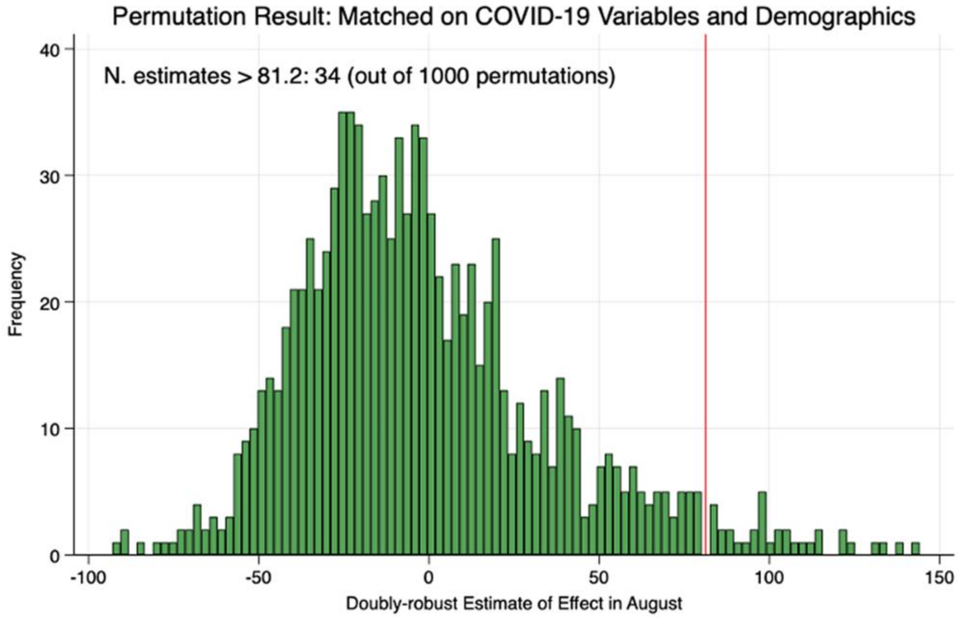

Fig. 5. ATE distributions for San Quentin State Prison connectivity permutation, August 2020. This figure displays DR estimates matched on pretransfer COVID-19 and demographic variables when permutating the connection between zip codes and San Quentin State Prison.

Corcoran and San Quentin, (ii) differences in the treatment-control match quality across the two prisons, and (iii) differences in the amount of staff-prisoner contact in Corcoran and San Quentin. First, zip codes connected to Corcoran had higher COVID-19 case rates prior to the prison outbreak in May 2020 compared to those connected to San Quentin. These preexisting higher COVID-19 case rates, combined with the nonlinear characteristics of COVID-19 infection, likely contributed to the amplified effect sizes observed postoutbreak in Corcoran. Second, the matched control group for Corcoran is less comparable to the treated group than in the San Quentin case, with a lower COVID-19 case rate in May before the outbreak started. This reduced matching quality may lead to an overestimation of the treatment effect, especially given the nonlinear COVID-19 dynamic. Third, in our smartphone data, we find that Corcoran is associated with higher contact levels (i.e. more time spent in Corcoran) among its staff compared to San Quentin. While Corcoran had robust structural mitigation measures, this high level of contact may offset some of these benefits, leading to higher overall contact rates and, consequently, larger observed effects on community transmission.

Discussion

Though the COVID-19 pandemic may be historically unique, the underlying epidemiology is not; like all upper respiratory infections, COVID-19 is easily transmitted across individuals who come into close contact with one another, and the period of infectious transmission frequently precedes the appearance of symptoms. In this article, we document geographic variation in how connected California communities are to prison environments and show evidence that these connections are relevant to the spread of infectious diseases from prisons to communities. We use high-quality, directly observed geographic movements of prison staff to and from their communities, while also illustrating how publicly available data from the Census Bureau identifies similar patterns of disease transmission. Our quasiexperimental setting means we can plausibly estimate the unidirectional vector of transmission from prison to surrounding communities. Of course, we focus only on short-term transmission; by construction, we exclude any subsequent transmission from a staff member's zip code to other zip codes. This means we estimate a lower bound of the total extent to which prisons accelerated the transmission of COVID-19 among the general population.

To provide context for the size of this potential source of upper respiratory tract infection transmission, we conducted a hypothetical, back-of-the-envelope calculation. Suppose that a new upper respiratory tract infection appeared in California prisons, with infection rates that were equal to COVID-19 in June 2020. Based on our zip code connectivity matrix and our estimated average treatment effects, we estimate that the prison staff vector would lead to a total of 14,724 community infections statewide 1 month later and 19,243 infections statewide 2 months later.[a]

Our analysis shows that disease outbreaks in prison have first-order impacts on the communities where prison staff live and spend time. Without negating the importance of understanding the implications of incarceration for the public health of the systematically disadvantaged communities that lose residents to prison, our research shows that prisons can also lead to negative public health consequences for the people who work in those facilities and their communities. This is an under-recognized vector of disease transmission that can be addressed through careful implementation of existing policies and recommended practices, including better enforcement of mask-wearing, routine staff testing, and paid medical leave for prison staff.

Materials and methods

Measuring prison-zip code connectivity

We measure prison-zip code connectivity by combining smartphone location data with prison facility boundary data from the DHS, with a focus on the 35 state prisons operating in California in 2020. Given the restrictions on smartphone usage within prison facilities, we expand the prison boundary to include neighboring geohash-7 (a 153 m × 153 m grid) areas covering the prison boundary. In other words, we define "expanded prison boundary" as all geohash-7 s covering both the prison fence line and adjacent spaces, which usually are parking lots (and sometimes highways, see Appendix Fig. S1). The probability of misidentification is minimal, as California state prisons are typically situated in rural areas and well-separated from other infrastructures.

To identify prison staff, we create a "daily presence filter" that allows us to exclude smartphones that are simply passing by or temporarily near a prison facility. For computational reasons, the first step is to identify all smartphones that are within our expanded prison boundary at any point. Next, we discretize the data into geohash-7 × half-hour intervals, requiring each smartphone spend at least 10 min in *the same geohash* cell during any 30-min window. This second filter ensures that we only classify phones belonging to individuals who genuinely stop at the prison, rather than those who may have entered the boundary momentarily. We validated this measurement in three ways. First, we compared the number of phones we identify with public sources, including LODES data (Fig. 1) and county-level staff counts from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW), which provides county-level counts of correctional officers as of June 2020. We find that the number of phones we identify is highly correlated with these data sources (Figs. 1 and S2). Second, to address the possibility of misclassification, we present results when excluding phones with only one ping in the prison (Table S14), which yields similar estimates to our main finding. Third, in Table S9 we demonstrate the robustness of our findings by varying the threshold for treated zip codes from 30-min to more than 10 h and find similar results across these specifications.

To identify connections between prison staff in San Quentin and zip codes, we link geohash-7 s to zip code tabulation areas (ZCTAs)—mailing areas covered by zip codes created by the Census Bureau and referred to as "zip codes" throughout the article—based on their centroid and define the phone's home zip code as the zip code that the phone spent the most time when outside of the expanded prison boundaries during June to October 2020. This approach ensures that transient visits, such as phones passing through a highway, do not lead to misclassification of zip codes as connected to prisons.

We replicate our analysis using the LODES data, which provides information detailing the number of jobs for each home–workplace census block pair. This allows us to identify the home location of prison staff and create an analogous connectivity measure between San Quentin State Prison and zip codes.[b] Specifically, we calculate the total primary jobs for individuals working in the block group encompassing San Quentin State Prison in 2020 and residing in different census tracts.[c] We then aggregate the tract-level job to the zip code level using a census tract-zip code crosswalk based on the proportion of the tract's population within each zip code. A zip code is similarly considered connected to a prison if it has any workers employed at the San Quentin State Prison.

Though LODES offers detailed spatial data on origin–destination workflow that enables research on disease transmission, it has limitations compared to smartphone location data. LODES data may reflect the administrative rather than the actual worksite location (29). It is released annually and may lack precision about the timing of geographic movements and is subject to confidentiality protections that may introduce noise or suppress information. Indeed, employees working in the block group where Corcoran is located are not reported in the LODES for this reason. Additionally, while less common in correctional settings than other industries, LODES does not differentiate between remote and "on site" work. All these data concerns could potentially complicate the inferences about the impact of connectivity between prisons and local communities. For the purposes of tracking disease, either the LODES or smartphone data could be more precise measures of worker connectivity, depending on the specific context.

Empirical strategy

We investigate the public health impact of prison-zip code connectivity by exploiting an exogenous COVID-19 outbreak in San Quentin State Prison in June 2020 due to a prisoner transfer from California Institution for Men. We examine whether the outbreak in San Quentin led to a differential increase in COVID-19 case rates—collected from the CDPH—in zip codes that were connected to San Quentin State Prison (i.e. treated zip codes) compared to those without this connection (i.e. control zip codes).[d]

We follow Callaway and Li (24) and estimate the effect of connecting to San Quentin State Prison after its outbreak in June 2020 using an unconfoundedness-type of strategy that compares connected zip codes with unconnected, matched zip codes that have similar pretransfer COVID-19 and similar demographic characteristics. This strategy is motivated by the fact that the spread of local COVID-19 cases is highly nonlinear over time. Common epidemiological models (e.g. Susceptible-Exposed-Infectious-Removed [SEIR], Susceptible-Infected-Recovered-Dead [SIRD]) treat COVID-19 transmission as a function of the number of currently infected individuals, the number of susceptible individuals in a location, and the transmission properties of COVID-19 (e.g. infection rate). Traditional difference-in-differences (DiD) strategies may introduce bias as different initial COVID-19 conditions between treated and control zip codes can lead to dynamic differences in how the pandemic evolves in different locations, violating the parallel trend assumption underlying the DiD strategy. Callaway and Li (24) propose an alternative identification strategy that compares locations with similar pretreatment "states" of COVID-19 case rates, suggesting this unconfoundedness-type of strategy is more suited to examine the nonlinearity of COVID-19 transmission.

Specifically, this strategy involves estimating the propensity score of being connected to San Quentin State Prison (i.e. inverse probability weighting [IPW]) and estimating changes in COVID-19 cases among unconnected zip codes (i.e. outcome regression [OR]), using the same set of covariates: pretransfer COVID-19 case rates and demographics from the 2015–2019 ACS aggregated to the zip code level. It provides a DR estimate of the effect of connecting to San Quentin prison on community COVID-19 transmission, in the sense that the estimate is consistent if either the propensity score model or OR model is correctly specified. To estimate the average treatment effect on the treated (ATT), we first use IPW by weighting the control zip codes based on the estimated propensity score, where we estimate the probability of being connected to the San Quentin conditional on pretransfer covariates:

$$P(tr = 1|X) = \Lambda(X\beta) \tag{1}$$

$$\mathrm{IPW} = \frac{\Lambda(X\beta)}{1 - \Lambda(X\beta)}, \tag{2}$$

where $\Lambda()$ represents the logistic function, X denotes the vector of pretransfer covariates. The ATT using the IPW approach is given by:

$$\mathrm{ATT}_{\mathrm{IPW}} = E(\Delta y(1)) - E(\Delta y(0);\, \mathrm{IPW}). \tag{3}$$

Next, we employ the OR approach, estimating changes in the outcome among the control zip codes conditional on the same set of covariates, and averaging over the distribution of covariates for treated zip codes to estimate the ATT:

$$\Delta y(0) = X_0\beta + e_i \tag{4}$$

$$\Delta\hat{y}(1) = X_1\hat{\beta} \tag{5}$$

$$\mathrm{ATT}_{\mathrm{OR}} = E(\Delta y(1)) - X_1\hat{\beta}. \tag{6}$$

Finally, we integrate both approaches in a DR estimator, which combines the strengths of IPW and OR:

$$\mathrm{ATT}_{\mathrm{DR}} = E(\Delta y(1) - X_1\hat{\beta}|tr = 1) - E(\Delta y(0) - X_0\hat{\beta}|tr = 0;\ \ \mathrm{IPW}). \tag{7}$$

The estimates under this strategy are consistent if either the model for the propensity score or OR is correctly specified (i.e. DR) (30). Moreover, these DR estimates of the impact of prison connectivity are compatible with epidemiological models that model the nonlinear spread of COVID-19 cases (24).

In our main analysis, we match treated and control zip codes on pretransfer COVID-19 variables, including the cubic polynomial functions of the number of cases in May 2020, cumulative cases in May 2020, and the logarithm of zip code population from 2015–2019 ACS, as well as demographic characteristics, including percentages of Black, White, and Hispanic residents; the percentage of residents with a college degree; median household income; and the percentage of residents above age 65, again from 2015–2019 ACS. Research shows that the transmission rate of COVID-19 could be influenced by neighborhood demographics including socioeconomic status such as income and education (31, 32), race/ethnicity (31–34), and age (35). Moreover, the extent to which observed case counts reflect the true number of infections is influenced by availability and access to tests, which are likely patterned by socioeconomic and demographic factors (36). Although our preferred specification includes both pretransfer COVID-19 and demographic characteristics for matching, the Appendix provides results when we match treated and control zip codes solely on COVID-19 variables. Appendix Figure S3 illustrates propensity score distributions before and after weighting for treated and control zip codes, showing nearly identical distribution for the propensity score among treated and control zip codes after the weighting. Table S2 also suggests that the pretransfer COVID-19 and demographic characteristics are balanced after the weighting adjustment.

Notably, our findings are quantitatively similar under alternative specifications. To account for variations in COVID-19 transmission rate by demographics, in Appendix Table S4, we additionally include interactions between pretransfer COVID-19 variables and demographic variables during matching in addition to the existing covariates. In Appendix Tables S5, S6, and S7, we find consistent estimates when excluding zip codes containing prisons or excluding zip codes that are linked to other two prisons (Avenal State Prison and Chuckawalla Valley State Prison) that had reported COVID-19 outbreaks during June 2020, mitigating concerns that the results are driven by these specific zip codes. In Appendix Fig. S4, we find similar estimates matched on pretransfer COVID-19 variables when permuting our measure of connectivity across zip codes and reestimating our ATE.

We also explore alternative definitions of treated groups. For instance, we define a connection in June 2020 using varying time duration cutoffs (i.e. zip codes are connected to a prison if phones from that zip code spend more than 0.5, 1, 1.5, 2, 2.5, 5, and 10 h in prisons) in Appendix Tables S8 and S9. In Appendix Table S10, we define treated groups based on preperiod connections from March 2020 to May 2020 instead of connections in June 2020. Appendix Tables S11 and S12 report estimates from alternative estimators in addition to the DR estimators, including the inverse propensity weighting estimator and the OR estimator. In Appendix Table S13, we replicate results when defining home zip codes using the zip code where the phone spent the most time from January to October 2020 (instead of the second half of 2020). In Appendix Table S14, we report results when we exclude phones with only one ping inside San Quentin State Prison in June 2020. Finally, in Appendix Table S15, we find a similar increase in case rates for zip codes connected to Corcoran State Prison compared to unconnected zip codes, which experienced a smaller internal COVID-19 outbreak. Across these analyses, our preferred strategy matches on both pretransfer COVID-19 and demographics (rather than COVID-19 case rates alone). Overall, the similarity of our estimates suggests that, in the summer of 2020, facility-level variation in prisoner protection from COVID-19 was greater than facility-level variation in protection for staff.

Notes

[a] Specifically, we define a prison as having experienced an outbreak if the number of new cases in that prison exceeds 100 in June 2020. Under this definition, 8 out of 35 state prisons experience a prison outbreak. We then estimate how these prison outbreaks will lead to increase in COVID-19 case rates in zip codes that are connected to prisons experiencing outbreak.

[b] Actual prison locations are reported in the LODES in California because of the structure of the California Department of Corrections and Rehabilitation (CDCR) payroll. This is not the case for government employees in all states.

[c] The boundary of the block group containing San Quentin State Prison closely matches the boundary of San Quentin itself.

[d] Technically, the COVID-19 case rate data are available at the zip code level, which we matched to ZCTAs using the zip code-ZCTA crosswalk provided by the department of Housing and Urban Development. In cases where a single ZCTA is associated with multiple zip codes (representing <2% of the observations), we use the maximum COVID-19 case rate among these zip codes.

Acknowledgments

The authors thank seminar participants at Columbia, Cornell, UCI, and UCLA for generous and insightful feedback. We would also like to thank the UCLA: Ziman Center for Real Estate, Price Center, and Center for Impact, for funding the data infrastructure critical to this project.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

The authors declare no external funding.

Author Contributions

Conceptualization: Y.Z., K.T., N.S., E.O., and M.K.C. Methodology: Y.Z., K.T., N.S., E.O., and M.K.C. Investigation: Y.Z. Visualization: Y.Z. and M.K.C. Funding: E.O. and M.K.C. Writing—original draft: Y.Z., K.T., N.S., and E.O. Writing—review & editing: Y.Z., K.T., N.S., E.O., and M.K.C.

Data Availability

All data used in this article are publicly available. ACS 5-year estimate and LEHD Origin–Destination Employment Statistics Data are available from the US Census Bureau (37, 38). Prison boundary data are available from DHS (39), QCEW are available from Bureau of Labor Statistics (40). Smartphone ping data are available by purchase from Veraset LLC (41). Replication materials containing all code required to reproduce our analyses and results are available at Harvard Dataverse: https://doi.org/10.7910/DVN/2PUGRA.

References

1 Healthy People. Social determinants of health: incarceration [accessed 2024 Aug 27]. https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/incarceration.

2 LeMasters K, *et al.* 2022. COVID-19 community spread and consequences for prison case rates. *PLoS One*. 17:e0266772.

3 Wallace D, *et al.* 2021. Is there a temporal relationship between COVID-19 infections among prison staff, incarcerated persons, and the larger community in the United States? *Int J Environ Res Public Health*. 18:6873.

4 Crawley E. 2002. Bringing it all back home? The impact of prison officers' work on their families. *Probation J*. 49:277–286.

5 Jiménez MC, *et al.* 2020. Epidemiology of COVID-19 among incarcerated individuals and staff in Massachusetts jails and prisons. *JAMA Netw Open*. 3:e2018851.

6 Wang EA, Zenilman J, Brinkley-Rubinstein L. 2020. Ethical considerations for COVID-19 vaccine trials in correctional facilities. *JAMA*. 324:1031–1032.

7 Saloner B, Parish K, Ward JA, DiLaura G, Dolovich S. 2020. COVID-19 cases and deaths in federal and state prisons. *JAMA*. 324:602–603.

8 Raifman J, *et al. COVID-19 U.S. State policy database*. Inter-University Consortium for Political and Social Research, Michigan, 2020.

9 Komarla A. In California jails, sheriffs have been left in charge of COVID mitigation. The results have been disastrous [accessed 2024 Aug 27]. https://uclacovidbehindbars.org/sheriffs-california-jails.

10 Office of the Inspector General California. Transfer of patients from CIM [accessed 2024 Aug 27]. https://www.oig.ca.gov/wp-content/uploads/2021/02/OIG-COVID-19-Review-Series-Part-3-E28093-Transfer-of-Patients-from-CIM.pdf

11 Nowotny KM, Seide K, Brinkley-Rubinstein L. 2021. Risk of COVID-19 infection among prison staff in the United States. *BMC Public Health*. 21:1036.

12 Toblin RL, Cohen SI, Hagan LM. 2021. SARS-CoV-2 infection among correctional staff in the federal bureau of prisons. *Am J Public Health*. 111:1164–1167.

13 Reinhart E, Chen DL. 2020. Incarceration and its disseminations: COVID-19 pandemic lessons from Chicago's Cook County Jail. *Health Aff*. 39:1412–1418.

14 Reinhart E, Chen DL. 2021. Association of jail decarceration and anticontagion policies with COVID-19 case growth rates in US counties. *JAMA Netw Open*. 4:e2123405.

15 Walter KS, *et al.* 2022. The impact of the COVID-19 pandemic on tuberculosis case detection and treatment outcomes in Brazil: a retrospective cohort study. *Lancet Reg Health Am*. 10:100220.

16 Weyant RB, *et al.* 2022. Simulating the impact of policy interventions on health outcomes: a decision-analytic model. *Med Decis Making*. 42:369–383.

17 Mabud TS, *et al.* 2019. Dynamic interactions of migration, poverty, and infection with multidrug-resistant tuberculosis. *PLoS Med*. 16:e1002730.

18 Lewis NM, *et al.* 2021. Community-associated outbreak of COVID-19 in a correctional facility. Utah, September 2020. January 2021. *MMWR Morb Mortal Wkly Rep*. 70:467–472.

19 Chen MK, Chevalier JA, Long EF. 2021. Nursing home staff networks and COVID-19. *Proc Natl Acad Sci U S A*. 118:e2015455118.

20 Chen MK, Christensen KL, John E, Owens E, Zhuo Y. 2023. Smartphone data reveal neighborhood-level racial disparities in police presence. *Rev Econ Stat*. 1–29.

21 Atkin D, Chen MK, Popov A. 2022. The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley. NBER Working Paper No. 30147.

22 Simes JT. 2021. *Punishing places: the geography of mass imprisonment*. University of California Press.

23 California Office of the Inspector General. COVID-19 Review Series. Part Three [accessed 2024 Aug 27]. https://www.oig.ca.gov/wp-content/uploads/2021/02/OIG-COVID-19-Review-Series-Part-3-E28093-Transfer-of-Patients-from-CIM.pdf.

24 Callaway B, Li T. 2023. Policy evaluation during a pandemic. *J Econometrics*. 236:105454.

25 Wright AL, Sonin K, Driscoll J, Wilson J. 2020. Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. *J Econ Behav Organ*. 180:544–554.

26 White A, Maloney E, Boehm M, Bleakley A, Langbaum J. 2022. Factors associated with COVID-19 masking behavior: an application of the health belief model. *Health Educ Res*. 37:452–465.

27 Yang L, *et al.* 2022. Sociocultural determinants of global mask-wearing behavior. *Proc Natl Acad Sci U S A*. 119:e2213525119.

28 Grossman G, Rexer JM, Thirumurthy H. 2020. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proc Natl Acad Sci U S A*. 117:24144–24153.

29 Graham MR, Kutzbach MJ, McKenzie B. *Design comparison of LODES and ACS commuting data products. Working papers 14-38*. Center for Economic Studies, U.S. Census Bureau, Washington, DC, 2014.

30 Sant'Anna PH, Zhao J. 2020. Doubly robust difference-in-differences estimators. *J Econom*. 219:101–122.

31 Magesh S, *et al.* 2021. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic review and meta-analysis. *JAMA Netw Open*. 4(11):e2134147–e2134147.

32 Samuel LJ, *et al.* 2021. Race, ethnicity, poverty and the social determinants of the coronavirus divide: US county-level disparities and risk factors. *BMC Public Health*. 21:1–11.

33 Abedi V, *et al.* 2021. Racial, economic, and health inequality and COVID-19 infection in the United States. *J Racial Ethn Health Disparities*. 8:732–742.

34 Mackey K, *et al.* 2021. Racial and ethnic disparities in COVID-19–related infections, hospitalizations, and deaths: a systematic review. *Ann Intern Med*. 174(3):362–373.

35 Pijls BG, *et al.* 2021. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open*. 11(1):e044640.

36 Hsu P, Hayes-Bautista DE. 2021. The epidemiology of diversity: COVID-19 case rate patterns in California. *J Immigr Minor Health*. 23(4):857–862.

37 [dataset] U.S. Census Bureau. 2023. *American community survey: 5-year estimates: comparison profiles 5-year*. U.S. Census Bureau.

38 [dataset] U.S. Census Bureau. 2023. *LEHD Origin-destination employment statistics data (2002-2021), version* 8.1. Washington (DC):

Longitudinal Employer-Household Dynamics Program, U.S. Census Bureau.
39 [dataset] Homeland Infrastructure Foundation Level Data (HIFLD). 2023. *Prison boundaries*. Washington (DC): U.S. Department of Homeland Security.
40 [dataset] Bureau of Labor Statistics (BLS). 2023. *Quarterly census of employment and wages*. Washington (DC): U.S. Department of Labor.
41 [dataset] Veraset, LLC. *Movement Dataset*. Veraset, LLC, Denver, CO, 2023. https://www.veraset.com/datasets/movement.