Incarceration and mortality in the United States

Elias Nosrati, Jacob Kang-Brown, Michael Ash,
Martin McKee, Michael Marmot, Lawrence P. King

The ongoing COVID-19 pandemic has spotlighted the role of America's overcrowded prisons as vectors of ill health, but robust analyses of the degree to which high rates of incarceration impact population-level health outcomes remain scarce. In this paper, we use county-level panel data and a novel instrumental variable technique to study the causal impact of penal expansion on communicable, non-communicable, and all-cause mortality rates. We find that a standard deviation increase in rates of imprisonment in one year causes 2.9 excess deaths from communicable diseases per 100,000 county population in the following year (95% CI: 2.1, 3.7; $p < 0.001$). We also find that higher incarceration leads to 22.6 excess deaths from non-communicable diseases (95% CI: 18.7, 26.5; $p < 0.001$) and 22.6 excess all-cause deaths (95% CI: 18.3, 26.9; $p < 0.001$) per 100,000 population. These findings are further corroborated by a between-unit analysis using coarsened exact matching and a simulation-based regression approach to predicting geographically anchored mortality differences.

Classifications: Social Sciences, Economic Sciences, Medical Sciences.
Key words: Prison incarceration, multi-cause mortality, instrumental variables.

Significance statement

The association between incarceration and ill health has been extensively described and debated, and the advent of COVID-19 has further spotlighted the role of America's overcrowded prisons as vectors of disease. However, rigorous causal estimates of how incarceration impacts mortality rates remain scarce, especially at the population level. This paper uses hitherto unavailable information on local counties over time and a compound instrumentation method to provide a comprehensive analysis of how incarceration affects geographically anchored patterns of mortality from both communicable and non-communicable diseases, but also from all causes combined. We find statistically robust and substantively important causal effects that operate both within and between counties over time and our analysis thus sheds novel light on the socio-spatial patterning of health inequality and its distal determinants at a high level of geographical resolution. We provide new empirical insights surrounding the nexus of punishment and population health which may inform future policy-making geared towards criminal justice reform and the elimination of avoidable mortality.

Introduction

The advent of COVID-19 has spotlighted the role of America's overcrowded prisons as vectors of ill health (Akiyama et al., 2020; Saloner et al., 2020). As of 16 October 2020, over 147,000 prisoners have contracted the disease and over 1,240 have died as a result. Amongst prison staff, corresponding figures are over 32,400 and 86, respectively (Marshall Project, 2020). However, prior to the ongoing pandemic, causal evidence of the link between high rates of incarceration and infectious disease mortality at the population level has been scarce. More generally, despite the historically unprecedented expansion of the American penal state since the 1970s, imprisonment has rarely been construed as a distal determinant of population health in its own right (Wildeman and Wang, 2017). In this paper, we use spatially disaggregated time-series data and a novel instrumental-variable approach to examine how local prison admission rates impact age-standardised death rates at the US county level. By drawing on extant scholarship on the health impacts of penal expansion, we hypothesise a causal association between high imprisonment rates and county-level mortality that is operant above and beyond the role of factors like income, education, or violent crime. Moreover, we hypothesise that high incarceration rates impact not only those who pass through the criminal justice system but also local populations at large. We provide a comprehensive analysis — to our knowledge, the first of its kind — of how incarceration affects geographically anchored patterns of mortality from communicable and non-communicable diseases and all causes combined.

Background and hypotheses

Since the early 1970s, the American penal state has undergone a historically unprecedented expansion. After 50 years of relative stability, the national jail and prison incarceration rate stood at 161 residents per 100,000 population in 1973. In 2007, the corresponding figure was 767. In absolute numbers rather than rates, this amounts to a shift from just under 400,000 to over 2.3 million individuals behind bars — a sevenfold increase in less than four decades (National Research Council, 2014). Beyond these aggregate numbers, imprisonment has emerged as a new stage in the life course of men of colour mired at the bottom of the class structure (Pettit and Western, 2004; Western, 2006; Wacquant, 2009). This is evidenced by how the cumulative risk of experiencing parental incarceration by age 14 amongst African American children born to high-school dropouts exceeds 50% (Wildeman, 2009).

A rich body of evidence has related penal expansion to declining health and deepening health inequality (for recent reviews, see Massoglia and Pridemore, 2015; Wildeman and Wang, 2017). In particular, overcrowded correctional facilities have been linked to infectious disease transmission (Massoglia, 2008; Ndeffo-Mbah et al., 2018) — a linkage that has been further spotlighted by the ongoing COVID-19 pandemic (Akiyama et al., 2020;

Saloner et al., 2020). Upon release from prison, former inmates experience mortality rates close to thirteenfold that of the comparable populace and are especially vulnerable during the first two weeks post-release, notably via acute stress-related psychosocial mechanisms (Binswanger et al., 2007; Zlodre and Fazel, 2012).

Moreover, previous scholarship has documented the ways in which high rates of incarceration act in cascading ways upon other social determinants, or "fundamental causes" (Link and Phelan, 1995), of health. Chief amongst such upstream determinants are the social and economic decay of neighbourhoods (Sampson and Loeffler, 2010), the disruption of social and family ties (Wildeman and Muller 2012; Wildeman, Schnittker, and Turney 2012), adverse childhood experiences related to parental incarceration (Wildeman, 2009; Turney, 2014), and enduring material deprivation and hardship in areas of concentrated disadvantage (Western, 2018). For instance, the incarceration of a family member has been shown to impair the wellbeing of non-incarcerated partners and children, notably due to declining household income, reduced parental investment, and unstable social relationships (Wildeman and Muller, 2012; Turney, 2014). At the community level, the criminal justice system plays a pivotal role in shaping the trajectories of neighbourhoods by removing prime-age men from their local communities, fragmenting family relations, and eroding social ties (Western, 2006; Sampson and Loeffler, 2010; National Research Council, 2014; Western, 2018).

Against this backdrop, we hypothesise that high rates of incarceration have a causal impact on a range of mortality outcomes, not only at the level of the individual but on a population scale. Drawing on the extant literature, our argument is that the experience of incarceration may prove deeply consequential not only for those who are incarcerated but also for their families, friends, and broader social connections (Gilmore, 2007; Wildeman and Wang, 2017; Wildeman and Muller, 2012). Thus the causal pathways from imprisonment to mortality most likely involve diverse modalities of "social sundering" (Therborn, 2013: 22–28) whereby the material and symbolic fabric of social life is eroded, for individuals and collectives alike (see also Wilkinson and Pickett, 2010; 2018). Given this plurality of pathways and mechanisms, we therefore expect the hypothesised causal effect to manifest in the form of mortality from both communicable and non-communicable diseases, within and between units of analysis over time.

Although previous studies have shed light on the effects of incarceration on health at the level of individuals and communities, robust evidence at the population level remains scarce, notably when it comes to the geographical patterning of different types of mortality rates. Our macroscopic approach allows us not only to generate a broad overview of how incarceration shapes population health, but also eschews the methodological challenges intrinsic to the use of individual-level survey data (however, see Daza, Palloni, and Jones, 2020), such as producing plausible causal identification strategies and constructing appropriate comparison groups (see Johnson and Easterling 2012; Wildeman, Wakefield, and Turney 2013). In what follows, we test our hypotheses by using county-level panel data

and a novel instrumentation technique suited to isolating exogenous treatment variation.

Data and methods

Outcome and control variables

The first of our three outcome variables is the county-level age-standardised mortality rate from communicable, maternal, neonatal, and nutritional diseases per 100,000 population between 1983 and 2014. To avoid terminological cluttering, we henceforth use "communicable" as a shorthand for "communicable, maternal, neonatal, and nutritional". The second outcome variable is mortality from all non-communicable diseases, and the third is all-cause mortality. All three of these variables are publicly available from the Institute for Health Metrics and Evaluation (IHME, 2017). Our treatment variable is the county-level annual prison admissions rate, generated by the Vera Institute of Justice using state corrections sources and the National Corrections Reporting Program by the Bureau of Justice Statistics which are converted into annual county-level rates per 100,000 residents aged 15–64 (Hinds et al., 2020). Six states — Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont — are excluded from the analysis due to lack of consistently collected prison admissions data. Due to certain discrepancies between our data sources in measuring county boundaries and accounting for changes to counties over time, the state of Virginia and a handful of counties from other states are also excluded from the final analysis. We employ a set of baseline control variables that are associated with both the treatment and the outcome, namely annual rates of violent crime, median household income, high school graduation rates, and the fraction of each county population that is African American, Hispanic, or any other non-White ethnic minority. These variables are all available from the US Census Bureau, except for the measure of violent crime which is extracted from the Federal Bureau of Investigation's Uniform Crime Reporting Program. Descriptive statistics are reported in Table 1.

Instrumental-variable models

To empirically assess how incarceration affects county-level mortality rates, we posit the following data-generating process:

$$Y_{it} = T_{i[t-k]}\beta + X_{it}\theta + \mu_i + \varphi_t + \varepsilon_{it}, \tag{1}$$

where Y_{it} denotes one of the three alternative outcome variables as measured in county i at time t; the treatment variable $T_{i[t-k]}$ is the county-level incarceration rate per 100,000 population, lagged by k years to allow for delayed effects; X is a vector of control variables; μ and φ capture unit- and time-fixed effects, respectively; and ε is a stochastic error term. Our principal quantity of interest is β, which is a causal effect parameter to be estimated. However, as visualised in Figure 1, we face a potential identification problem wherein the

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Mortality from communicable disease	65,237	50	13	15	263
Mortality from non-communicable disease	65,237	841	115	247	1,499
All-cause mortality	65,237	972	140	323	1,832
Incarceration rate per 100,000 population	65,237	268	205	0.0	2,583
Violent crime rate per 100,000 population	65,237	278	267	0.0	5,972
Median household income ($)	65,237	47,105	11,709	17,583	125,705
High school graduation rate	65,237	0.8	0.1	0.3	1.0
Fraction African Americans	65,237	0.1	0.1	0.0	0.9
Fraction Hispanics	65,237	0.1	0.1	0.0	1.0
Fraction other ethnic minority	65,237	0.02	0.1	0.0	0.9

Notes: All variables, listed in the first column, are measured at the county level. The second column lists the number of observed county-years. The three outcome variables — communicable, non-communicable, and all-cause mortality rates per 100,000 population — are taken from the Institute for Health Metrics and Evaluation US Health Map database (IHME, 2017). The incarceration rate is per 100,000 population aged 16–64 and is constructed by the Vera Institute of Justice (Hinds et al., 2020). The measure of violent crime is extracted from the Federal Bureau of Investigation's Uniform Crime Reporting Program. All remaining variables are taken from the US Census Bureau.

estimated relation between the treatment variable T and the outcome variable Y is biased by some unmeasured confounder U, even after controlling for observed covariates X. In our case, U might denote unobserved variables that simultaneously affect incarceration and mortality, such as locally contingent healthcare or welfare-related policy shocks.

To address this concern, we construct an instrumental variable, Z, that is suited to isolating exogenous variation in T. To do this, we adopt a compound instrument derived from the interaction between the county-specific average exposure to incarceration over the sample period and annual nationwide per capita correctional spending. This instrument meets the relevance criterion insofar as increasing correctional expenditure is predictive of higher rates of incarceration. It also meets the exclusion criterion insofar as annual aggregate correctional spending is independent of any given county, to the effect that unit-specific shocks that deviate from a county's long-run average exposure to imprisonment result from a treatment assignment mechanism that is orthogonal to that county's potential outcomes. In other words, the outcome of interest in units with varying propensities to incarcerate will not be affected by changes in aggregate correctional spending other than through the impact of incarceration.

We believe that this proposed instrumentation method constitutes an advance in the study of the incarceration-health nexus. A recent study by Weidner and Schultz (2019) uses a cross-sectional design in which correctional spending alone is used as an instrumental variable. We argue that the methodological framework of the present paper provides a more stringent framework for causal inference by virtue of the time-series dimension of our data. Not only are year- and unit-specific attributes netted out by de-meaning through entities, but lagged effects are also incorporated into our model design. The two-way fixed-effects model thus constitutes a rigorous approach that eliminates any confounders that either remain stable over time — such as county- or state-level institutional factors — or form part of any aggregate time trends, whilst also allowing for dynamic relationships. This combination of factors leads us to believe that we are better positioned to isolate exogenous shocks that operate above and beyond individual units' default exposure to incarceration.

We thus obtain an instrument $Z_{it} = \overline{T}_i \times C_t$, where $\overline{T}_i$ is county i's average incarceration rate over the sample period and C_t is the aggregate per capita expenditure on the construction and maintenance of correctional facilities across all states in year t. The latter variable is obtained from the Bureau of Justice Statistics' Justice Expenditure and Employment Series and is measured every few years. A spline function is then applied to impute missing values through interpolation between observed years, the result of which is visualised in SI Figure A1. Our two-stage regression model now has the following selection equation:

$$T_{it} = Z_{it}\tau + X_{it}\eta + \alpha_i + \delta_t + \upsilon_{it}. \tag{2}$$

We then re-specify the model in equation (1) as follows, with $\widehat{T}$ being a vector of fitted

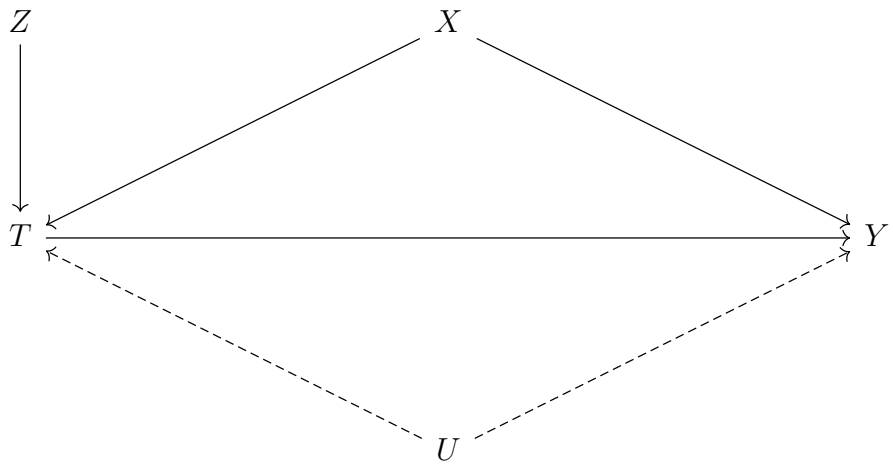

Figure 1: Causal graph depicting the effect of the treatment variable (T = incarceration) on the outcome (Y = life expectancy or premature mortality risk), identified via a compound instrument (Z = unit-specific average incarceration rate $\times$ annual nationwide correctional spending), net of both measured covariates ($X = \{$violent crime, household income, high school education, demographics$\}$) and unmeasured confounders (U).

values from equation (2):

$$Y_{it} = \widehat{T}_{i[t-k]}\beta + X_{it}\theta + \mu_i + \varphi_t + \varepsilon_{it}. \tag{3}$$

We set $k = 1$ for communicable disease mortality and $k = 3$ for the other two outcomes to allow for temporally heterogeneous aetiological windows, as we expect the impact of incarceration to take longer to manifest at the population level for chronic disease and all-cause mortality. To empirically assess the strength of the chosen instrument, we compare the model in equation (2) to a restricted first-stage regression in which the effect τ of Z on T is set to be null, obtaining a χ^2 test statistic of 8,057 on 1 degree of freedom ($p < 0.001$). Hence Z comfortably satisfies the benchmark for identifying a strong instrument. Our model accounts for (a) any time-invariant confounders, even if these are unobserved, by isolating variation within counties over time and (b) any aggregate trends that affect all counties simultaneously.

Matching and between-county models

We complement our analysis of within-county variation over time with a model of long-run mortality differences between counties by averaging across units over our sample period. Despite the fact that fixed-effects models are typically preferred in causal inference and that between-unit variation rarely yields plausible estimates of a causal relationships, we are nevertheless interested in the between-county variation because a sole focus on within-county variation over time prevents us from examining a key quantity of interest, namely the magnitude of disparity in mortality burdens between counties. However, in order to render the corresponding parameter estimates more plausible, we employ matching as a non-parametric form of pre-processing the data (Ho et al., 2007; Iacus et al., 2018).

The goal of matching is to reduce inefficiency, bias, and model dependence. It is a non-model-based approach to preparing the data for parametric analysis with a view to mimicking experimental research designs. In non-technical terms, matching seeks to select units of analysis (counties) that are similar if not identical to one another in all respects except for one: whether or not they are exposed to a key variable of interest. In the present case, the quantity of interest is the effect of high rates of incarceration on mortality rates, above and beyond the endogenous associations between incarceration and factors like income, education, or crime. Applying a matching algorithm will help "match" counties that share key characteristics, with the exception that some have high incarceration rates and others have low incarceration rates. This will facilitate a more precise account of the link between penal expansion and the local mortality burden. In more technical terms, let Y_i designate the outcome variable of interest (mortality), let $T_i \in \{0, 1\}$ designate a dichotomous treatment variable (low versus high incarceration rates), and let X_i designate a series of pre-treatment covariates (income, education, crime, demographic composition, etc.). Then the treatment effect β on a treated unit i is $\beta = Y_i(T_i = 1) - Y_i(T_i = 0)$.

However, the last term of this equation, $Y_i(T_i = 0)$, is an unobserved counterfactual. One can estimate this quantity with Y_j from control units (indexed by j) that are matched on relevant covariates (i.e., $X_i \approx X_j$) such that the estimated counterfactual quantity, $\widehat{Y_i}(T_i = 0)$, is equal to $Y_j(T_j = 0)$. Unmatched units are pruned from the data set to improve empirical covariate balance between treatment and control groups in the sample, and the parametric model is applied to the pruned rather than to the raw data. As a result, the functional form of the parametric specification is subject to less arbitrary model dependence.

In the analysis below, we employ what is known as coarsened exact matching. This form of matching proceeds as follows. For lack of being able to match on *exact* values of continuous covariates, this algorithm temporarily "coarsens" the covariates X into sub-categories (defined via a non-parametric histogram estimator). It then applies exact matching on the coarsened X, $c(X)$, before sorting observations into strata, each with unique values of $c(X)$. Any stratum with zero treated or control units is pruned from the data set. The algorithm then passes the original (uncoarsened) units — except for the pruned ones — on to the matched data set that is used in the parametric analysis (for further details, see Iacus et al. [2012]).

After obtaining a matched data set, we regress Y on T alone using simple ordinary least squares regression, since covariate balance is obtained through matching. We then adopt a simulation-based approach to presenting key quantities of interest (King, Tomz, and Wittenberg, 2000). We collect our model estimates in the stacked column vector $\hat{\psi} = \{\hat{\beta}, \hat{\sigma}^2\}$, which forms the mean of a multivariate Normal distribution with variance equal to the model covariance matrix, $\hat{\mathbb{V}}(\hat{\psi})$. We reduce model dependence by drawing tens of thousands of numbers from this distribution and averaging uncertainty across the simulated parameter estimates. This further allows us to simulate counterfactuals by comparing expected values of each of the outcome variables across treatment states, with $T = 0$ for counties below mean exposure and $T = 1$ for counties with above mean exposure to incarceration. The simulated expected values are used to visualise uncertainty surrounding model parameters and to directly compare the distributions of $\mathbb{E}(Y \mid T = 0)$ and $\mathbb{E}(Y \mid T = 1)$.

Sensitivity analyses

Given the lack of instrumentation in this cross-sectional setting, we refrain making causal claims. However, we conduct a simple non-parametric sensitivity analysis that allows us to precisely quantify the amount of unmeasured confounding that would in theory be required to eliminate our estimated treatment effect $\hat{\beta}$. For (theoretically dichotomised) treatment and control units, let U denote an unmeasured confounder. Then the bias factor, $\mathcal{B}$, is defined as the difference between $\hat{\beta}$ and what $\hat{\beta}$ would have been had we controlled for U as well, net of our other control variables. We assume that U is binary

and that the effect of U on Y is the same across both treatment states (i.e., no U-by-T interaction). We then define

$$\gamma = \mathbb{E}(Y \mid U = 1,\ T,\ X) - \mathbb{E}(Y \mid U = 0,\ T,\ X)$$

as the effect of the unmeasured confounder on the outcome, net of the treatment and control variables. We also define

$$\delta = \mathbb{P}(U = 1 \mid T = 1,\ X) - \mathbb{P}(U = 1 \mid T = 0,\ X)$$

as the difference in the prevalence of the unmeasured confounder between the treatment and control groups. Then it can be shown that $\mathcal{B} = \gamma \times \delta$ (VanderWeele and Arah, 2011; VanderWeele, 2015: 68–69). In assessing the sensitivity of our model coefficients to unmeasured confounding, we ask how large γ would have to be in order to reduce our estimated effect size $\hat{\beta}$ to zero. We address this question by visualising how $\mathcal{B}$ changes as the two sensitivity parameters (co-)vary across a range of possible values.

As a final robustness check, we run a series of cross-sectional regressions with data from 2014 alone, without matching. For this particular year, we have access to additional control variables that help inform the sensitivity analysis, including residential segregation by race, unemployment and poverty rates, and intergenerational income mobility. These additional data and their sources are described in SI Table A1. Due to issues of multicollinearity, we present a series of regression models in which the control variables are added and removed one at a time. We then assess how the coefficient for incarceration changes in response to each new covariate. All analyses are conducted in R version 4.0.2 (R Core Team, 2020).

Findings

Panel data regressions

Table 2 shows results for a set of two-way fixed-effects regressions wherein the incarceration variable is instrumented as described above. We find, as shown in the second column, that a standard deviation increase in rates of imprisonment in one year causes 2.9 excess deaths from communicable diseases per 100,000 county population in the following year (95% CI: 2.1, 3.7; $p < 0.001$). Expressed as a percentage change, this amounts to a 7.2% increase in the local communicable death rate — a substantively large effect size. Now lagging the treatment by three years to allow for medium-run effects to manifest, we find, as shown in the third column, that higher incarceration leads to 22.6 excess deaths from non-communicable diseases (95% CI: 18.7, 26.5; $p < 0.001$), equivalent to a 3.4% increase in such deaths. Finally, as shown in the fourth column, the treatment effect for mortality from all causes is 22.6 excess deaths (95% CI: 18.3, 26.9; $p < 0.001$), or a 3.3% increase.

Our models also indicate, as expected, that increasing median household income reduces all three categories of cause of death, whereas rising violent crime has an adverse impact. The remaining coefficients in the fixed-effects model must be interpreted with caution given that within-county variation in the fraction of high school graduates or changes in the ethno-racial composition are typically small and are nearly time-invariant. For instance, we see that a higher fraction of high school graduates is associated with higher mortality rates. We interpret this as a result of either insufficient within-unit variance or within-unit regression to the mean that is conducive to a spurious association. Our suspicion appears to be confirmed by running a separate cross-sectional regression (not displayed) in which the sign of the education variable reverses. Similar caution applies to interpreting changes in the ethno-racial composition of counties, which remains stable for most of the sample period and is also skewed due to high levels of segregation. Alternative model specifications that either transform or remove the baseline control variables make little to no substantive difference to the estimated treatment effects of incarceration.

For the sake of comparison, we also run non-instrumented versions of the two-way fixed-effects models. As shown in SI Table A2, these consistently produce smaller parameter estimates, but remain highly robust. We surmise that the discrepancy in effect sizes derives from attenuation bias in the non-instrumented panel regression — possibly due to measurement error or omitted variable bias — or from differences between the *local* average treatment effect estimated by the instrumented models and the *population* average treatment effect estimated by the non-instrumented models (see Card, 2001).

Between-county matched regressions

We proceed to the between-county analysis by applying a matching algorithm to time-averaged versions of all our covariates after splitting counties into those with above versus below mean exposure to incarceration. The matching procedure results in a pruned data set composed of 1,694 counties. As reported in SI Table A3, the diagnostics reveal a high degree of balance improvement since the empirical covariate distributions in both the treatment and control groups are now similar, meaning the smaller sample size strengthens rather than undermines the subsequent statistical inference. We then regress our three outcome variables on the treatment variable using simple ordinary least squares, the results of which are displayed in Table 3. A standard deviation increase in incarceration rates are associated with 4.3 excess communicable deaths (95% CI: 3.7, 4.9; $p < 0.001$), 44.2 excess non-communicable deaths (95% CI: 39.7, 48.7; $p < 0.001$), and 56.1 excess all-cause deaths (95% CI: 50.8, 61.4; $p < 0.001$). Expressed in terms of semi-elasticities, this amounts to a 8.9%, 5.3%, and 5.9% increase in mortality, respectively.

To get a more intuitive sense of what these numbers mean in substantive terms, we predict and plot the conditional expectation of each outcome variable by treatment status using a simulation-based approach, as described in the Data and methods section.

Table 2: Instrumented two-way fixed-effects regression models

	Communicable	Non-Communicable	All-Cause
Incarceration rate $(t-1)$	2.9**		
	(0.4)		
Incarceration rate $(t-3)$		22.6**	22.6**
		(2.0)	(2.2)
Violent crime rate	0.7**	4.3**	5.7**
	(0.1)	(0.7)	(0.8)
Median household income	−2.6**	−14.0**	−16.1**
	(0.3)	(1.5)	(1.7)
High school graduation rate	4.4**	39.7**	46.0**
	(0.3)	(2.5)	(2.8)
Fraction African Americans	7.4**	7.8	20.5**
	(0.8)	(5.2)	(6.1)
Fraction Hispanics	1.2	17.0**	12.4*
	(0.5)	(3.2)	(3.8)
Fraction other ethnic minority	−4.5**	−25.2**	−33.6**
	(0.8)	(5.5)	(6.6)
Wald test	592**	788**	793**
Observations	63,135	58,435	58,435

Notes: The outcome variables are age-standardised mortality rates from communicable diseases in the second column, from non-communicable diseases in the third column, and from all causes in the fourth column. The incarceration variable, lagged by one and three years, is instrumented as described in the Data and methods section. The corresponding parameter estimates are interpreted as the excess number of deaths per 100,000 county population caused by a standard deviation increase in incarceration rates. Robust standard errors are shown in parentheses below each parameter estimate. Statistical significance levels: $^*p < 0.01$; $^{**}p < 0.001$.

Table 3: Between-county matched regression models

	Communicable	Non-Communicable	All-Cause
Incarceration rate	4.3	44.2	56.1
	(0.3)	(2.3)	(2.7)
Multiple R^2	14.8%	18.4%	20.2%
Observations	1,694	1,694	1,694

Notes: The outcome variables are age-standardised mortality rates from communicable diseases in the second column, from non-communicable diseases in the third column, and from all causes in the fourth column. The association between treatment and outcome is estimated by applying a simple linear regression model to a pruned data set that is pre-processed using coarsened exact matching. Counties are matched on the variables listed in the Data and methods section (see also SI Table A3). All variables are time-averaged over the sample period. Parameter estimates are interpreted as the number of excess deaths associated with a standard deviation increase in incarceration rates. Standard errors are shown in parentheses below each parameter estimate. All parameter estimates are statistically significant at $p < 0.001$.

FIGURE 2 shows that in counties with low rates of incarceration, mortality from communicable diseases is expected to be 44.4 deaths per 100,000 population (95% CI: 43.8–45.0; $p < 0.001$). In counties with high rates of incarceration, the corresponding number is 52.0 (95% CI: 51.4–52.7; $p < 0.001$). For non-communicable diseases, a shift from control to treatment increases the expected mortality rate from around 800 (95% CI: 795–806; $p < 0.001$) to around 879 deaths per 100,000 population (95% CI: 873–884; $p < 0.001$). For all-cause mortality, the corresponding numbers are 917 (95% CI: 910–923; $p < 0.001$) and 1,016 (95% CI: 1,009–1,022; $p < 0.001$), respectively.

FIGURE 3 visualises the sensitivity analysis using the parameter estimates from TABLE 3, with δ denoting the degree of selection on the unmeasured confounder across the two treatment states (ranging from 0 to 1, with higher values indicating a higher prevalence of the confounder in the treatment group — i.e., in counties with high rates of incarceration), and γ denoting the magnitude of the effect of U on the outcome, above and beyond that of the treatment and other controls, that would be required to completely eliminate the effect of incarceration on the three outcome variables. The reader will note that even for unusually high levels of selection on the unmeasured confounder, the effect of U on the outcome would have to be large in order to nullify that of incarceration, especially for non-communicable disease deaths and all-cause mortality. For instance, even when the difference in the prevalence of the confounder between the treatment and control groups is as high as 0.8 — an unlikely scenario — U would have to cause around half a dozen excess communicable deaths, over 50 non-communicable deaths, and almost 75 all-cause deaths per 100,000 population — above and beyond the effects of T and X — to eliminate our model estimates. We believe it is plausible that most relevant confounders are already included in our matrix of covariates. As such, a more plausible value of δ would be at the lower end of the X-axis in Figure 3. At, say, $\delta = 0.2$, the effect of U on Y would have to be nearly 25 excess communicable deaths, around 225 non-communicable deaths, and around 275 all-cause deaths, which seems highly improbable. In short, an unusually large amount of unmeasured confounding would be needed to explain away the estimated impact of incarceration on between-county mortality patterns.

Cross-sectional regressions

Finally, we assess the robustness of the hypothesised relation between incarceration and mortality by running a series of cross-sectional regressions with additional data from 2014, without pre-processing the data via matching. Additional control variables include the local unemployment and poverty rates, a measure of absolute income mobility at the county level, income inequality as measured by a local Gini index, residential segregation by race, and the percentage of the county population with no health insurance. To avoid over-specification, we add and remove one control variable at a time. However, we adjust for state-fixed effects in all models. As reported in SI TABLES A4–A6, the estimated

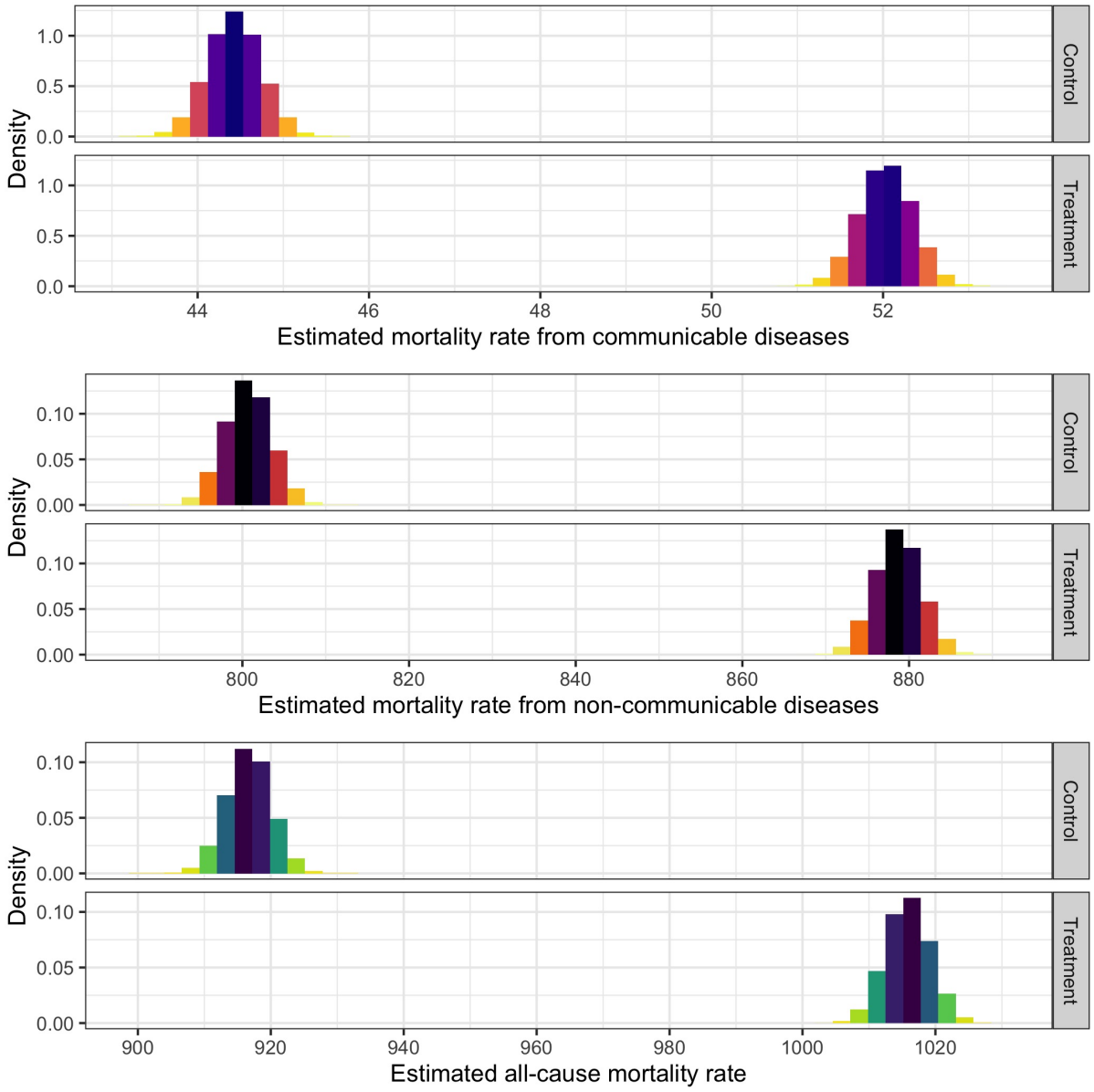

Figure 2: Density plots of expected outcome values conditional on treatment state. In the top panel, the outcome variable is mortality from communicable disease, in the middle panel the outcome variable is mortality from non-communicable diseases, and in the bottom panel the outcome variable is all-cause mortality per 100,000 population. Each model compares counties with incarceration rates at one standard deviation below the mean ("Control") to those with incarceration rates at one standard deviation above the mean ("Treatment"). The association between treatment and outcome is estimated by applying a simple linear regression model to a pruned data set that is pre-processed using coarsened exact matching. Counties are matched on the variables listed in the Data and methods section (see also SI Table A3). All variables are time-averaged over the sample period. $N = 1,694$.

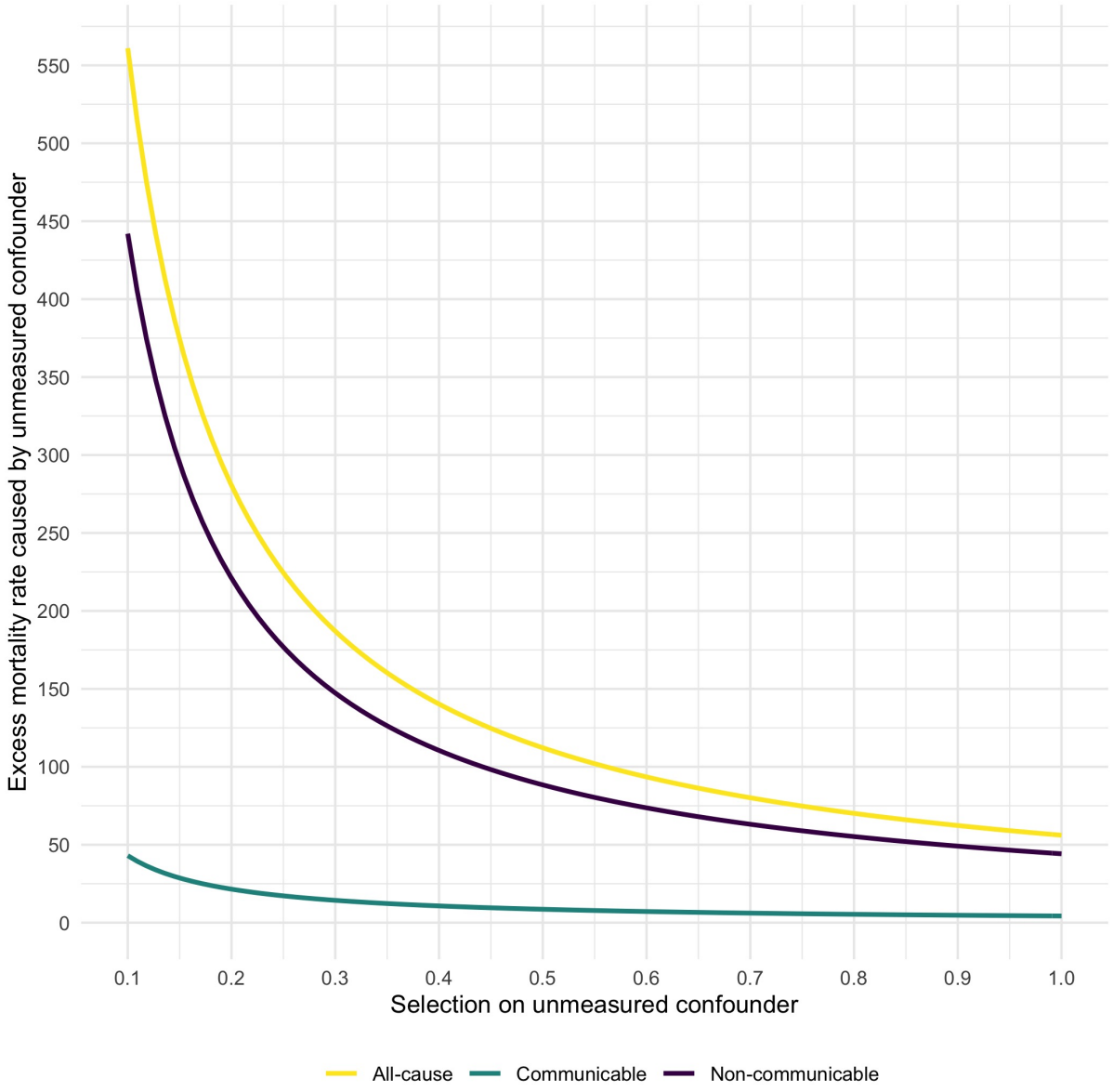

Figure 3: Sensitivity analysis plot to assess unmeasured confounding of the estimated effect $\hat{\beta}$ of incarceration on each of the outcomes in Table 3. Values of δ (X-axis) and γ (Y-axis) that lie on the lines would completely eliminate the corresponding effect estimates. Values above the plotted curve would reverse the sign of the estimated effect.

association between incarceration and each of the outcome variables remains remarkably stable across all specifications, which further confirms the robustness of our principal findings.

Concluding discussion

Our findings confirm the hypothesis that high rates of incarceration exert a substantively large impact on county-level mortality rates. Our joint usage of variation within and between units demonstrates that penal expansion can be construed as a distal determinant of declining health and deepening health inequality across the United States. However, we acknowledge the limitations of this study. First of all, for lack of being able to empirically verify that our models capture exogenous treatment variation, we recognise that our parameter estimates may suffer from residual confounding or other sources of bias. Nevertheless, in the panel data analysis, the use of a novel instrumentation technique coupled with unit- and time-fixed effects provides a stringent framework for causal inference that minimises the likelihood of obtaining spurious associations. In the cross-sectional analysis, we have sought to adjust for the most important and likely confounders of the relevant relationships and we have used matching as a means of mimicking an experimental research design. Both analyses yield robust parameter estimates and are not subject to high levels of model dependence. Our sensitivity analysis suggests that an inordinate amount of unmeasured confounding would be required to explain away the estimated effect of incarceration on mortality outcomes.

Second, although our findings provide meaningful quantitative estimates of the hypothesised causal associations, we do not have the data to flesh out the relevant pathways or to detail the precise mechanisms by which high rates of incarceration exert the kinds of effects that we propose. We are also not able to capture the broader correctional population, such as those on probation or parole, nor potentially relevant patterns of migration or other demographic changes. Moreover, our data do not allow us to estimate how much of the excess mortality is due to high death rates amongst former prisoners and how much is due to spillover effects on local areas. Future research should seek to integrate multilevel data that account for the complex interconnections between individuals, neighbourhoods, local communities, and the criminal justice system.

Third, we estimate an average treatment effect, yet existing research on incarceration shows that the penal state disproportionately targets a specific fraction of the American population, namely African Americans at the bottom of the social structure (Wacquant, 2009). With our current data and ecological approach, we are unable to assess any potential treatment effect heterogeneity — that is, whether incarceration is more harmful to some than to others. We are also unable to offer a more refined treatment of social and economic factors such as income, education, or ethno-racial division, all of which are imperfectly measured at an aggregate level in our data set.

These limitations do not prevent us from concluding that high rates of incarceration shape unequal life chances in the United States and can harm population health. We provide evidence for a robust and substantively large net causal linkage between incarceration and communicable, non-communicable, and all-cause mortality rates. This implies that protective rather than punitive criminal justice policies may help to shield vulnerable populations from premature mortality and to reduce regional inequalities in health and wellbeing.

Bibliography

Akiyama MJ, Spaulding AC, Rich JD. Flattening the curve for incarcerated populations — Covid-19 in jails and prisons. New England Journal of Medicine. 382(22): 2075–2077.

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, Koepsell TD. 2007. Release from Prison — A High Risk of Death for Former Inmates. *New England Journal of Medicine*. 356(2): 157–165.

Card D. 2001. Estimating the Return to Schooling. *Econometrica*. 1127–1152.

Daza S, Palloni A, Jones, J. 2020. The Consequences of Incarceration for Mortality in the United States. *Demography*. 57: 577–598.

Gilmore, RW. 2007. *Golden Gulag: Prisons, Surplus, Crisis, and Opposition in Globalizing California*. Berkeley: University of California Press.

Hinds O, Lu O, Wallace-Lee J, Kang-Brown J. 2020. *Reconstructing How Counties Contribute to State Prisons*. New York: Vera Institute of Justice.

Ho DE, Imai K, King G, Stuart EA. 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*. 15: 199–236.

Iacus SM, King G, Porro G. 2019. A Theory of Statistical Inference for Matching Methods in Causal Research. *Political Analysis*. 27(1): 46–68.

Iacus SM, King G, Porro G. 2012. Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis*. 20(1): 1–24.

IHME. 2017. *United States Life Expectancy and Age-specific Mortality Risk by County 1980–2014*. Seattle, United States: Institute for Health Metrics and Evaluation.

Johnson EI, Easterling B. 2012. Understanding Unique Effects of Parental Incarceration on Children: Challenges, Progress, and Recommendations. *Journal of Marriage and Family* 74(2): 342–356.

King G, Tomz M, Wittenberg J. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*. 44: 341–355.

Link BG, Phelan J. 1995. Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*. Special issue: 80–94.

Marshall Project. 2020. A State-by-State Look at Coronavirus in Prisons. https://tinyurl.com/y8hv3kc5. Accessed 16 October 2020.

Massoglia M. 2008. Incarceration as exposure: the prison, infectious disease, and other stress-related illnesses. Journal of Health and Social Behavior. 49(1): 56–71.

Massoglia M, Pridemore WA. 2015. Incarceration and Health. *Annual Review of Sociology*. 41: 291–310.

National Research Council. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences*. Edited by Travis J, Western B, Redburn S. Washington, DC: National Academies Press.

Ndeffo-Mbah ML, Vigliotti VS, Skrip LA, Dolan K, Galvani AP. 2018. Dynamic models of infectious disease transmission in prisons and the general population. *Epidemiologic Reviews*. 40(1): 40–57.

Opportunity Insights, n.d. https://opportunityinsights.org. Accessed 12 September 2020.

Pettit B, Western B. 2004. Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration. *American Sociological Review* 69(2): 151–169.

R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Saloner B, Parish K, Ward JA, DiLaura G, Dolovich S. 2020. COVID-19 Cases and Deaths in Federal and State Prisons. *Journal of the American Medical Association*. 324(6): 602–603.

Sampson RJ, Loeflller C. 2010. Punishment's place: the local concentration of mass incarceration. *Daedalus*. 139(3): 20–31.

Therborn, G. 2013. *The Killing Fields of Inequality*. Cambridge: Polity Press.

Turney K. 2014. Stress proliferation across generations? Examining the relationship between parental incarceration and childhood health. *Journal of Health and Social Behavior*. 55(3): 302–319.

VanderWeele TJ. 2015. *Explanation in Causal Inference. Methods for Mediation and Interaction*. New York: Oxford University Press.

Vanderweele TJ, Onyebuchi AA. 2011. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22(1): 42–52.

Wacquant L. 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Durham: Duke University Press.

Wildeman C. 2009. Parental imprisonment, the prison boom, and the concentration of childhood disadvantage. *Demography*. 46(2):265–280.

Wildeman C, Wang EA. 2017. Mass incarceration, public health, and widening inequality in the USA. *The Lancet*. 389(10077): 1464–1474.

Wildeman C, Muller C. 2012. Mass imprisonment and inequality in health and family life. *Annual Review of Law and Social Science*. 8: 11–30.

Wildeman C, Wakefield S, Turney J. 2013. Misidentifying the effects of parental incarceration? A comment on Johnson and Easterling (2012). *Journal of Marriage and Family* 75(1): 252–258.

Wildeman C, Schnittker J, Turney K. 2012. Despair by Association? The Mental Health of Mothers with Children by Recently Incarcerated Fathers. *American Sociological Review*. 77: 216–243.

Wilkinson R, Pickett K. 2018. *The Inner Level: How More Equal Societies Reduce Stress, Restore Sanity and Improve Everyone's Well-being*. London: Penguin.

Wilkinson R, Pickett K. 2010. *The Spirit Level: Why Equality is Better for Everyone.* London: Penguin.

Weidner RR, Schultz J. 2019. Examining the relationship between US incarceration rates and population health at the county level. *SSM Population Health* 9: 100466.

Western B. 2018. *Homeward: Life in the Year After Prison.* New York, NY: Russell Sage Foundation.

Western B. 2006. *Punishment and Inequality in America.* New York, NY: Russell Sage Foundation.

Zlodre J, Fazel S. 2012. All-Cause and External Mortality in Released Prisoners: Systematic Review and Meta-Analysis. *American Journal of Public Health.* 102(12): e67–e75.

Supplementary Information (SI)

Table A1: Descriptive statistics: cross-sectional data from 2014

Statistic	N	Mean	St. Dev.	Min	Max
Unemployment rate	2,830	6.3	2.2	1.2	24.0
Poverty rate	2,830	0.1	0.1	0.02	0.5
Absolute income mobility	2,673	47.5	6.1	25.1	68.3
Income inequality	2,830	0.4	0.1	0.2	0.8
Racial segregation	2,830	0.1	0.1	0.0	0.7
Percentage uninsured	2,830	18.4	5.5	3.6	41.4

Notes: All variables, listed in the first column, are measured at the county level in the year 2014. The second column lists the number of observations. The unemployment rate variable is taken from the US Bureau of Labor Statistics. All other variables are from the Opportunity Insights database (n.d.). Absolute income mobility is measured as the county-level percentage of children who earn more than their parents in 2014. Income inequality is measured as the county-level Gini index within the bottom 99% of the income distribution. Racial segregation is measured as residential segregation by race. The final variable measures the percentage of the county population without health insurance.

Table A2: Non-instrumented two-way fixed-effects regression models

	Communicable	Non-Communicable	All-Cause
Incarceration rate ($t-1$)	0.8**		
	(0.1)		
Incarceration rate ($t-3$)		8.6**	10.5**
		(0.6)	(0.6)
Violent crime rate	0.8**	4.9**	6.1**
	(0.1)	(0.6)	(0.8)
Median household income	−2.7**	−13.9**	−16.0**
	(0.3)	(1.5)	(1.8)
High school graduation rate	4.9**	43.1**	49.0**
	(0.3)	(2.5)	(2.8)
Fraction African Americans	8.0**	12.1	24.2**
	(0.8)	(5.4)	(6.3)
Fraction Hispanics	1.3*	19.2**	14.3**
	(0.5)	(3.2)	(3.8)
Fraction other ethnic minority	−5.1**	−27.8**	−35.9**
	(0.8)	(5.4)	(6.5)
F test	2,149**	1,822**	2,189**
Observations	63,135	58,435	58,435

Notes: The outcome variables are age-standardised mortality rates from communicable diseases in the second column, from non-communicable diseases in the third column, and from all causes in the fourth column. The corresponding parameter estimates are interpreted as the excess number of deaths per 100,000 county population associated with a standard deviation increase in each predictor. Robust standard errors are shown in parentheses below each parameter estimate. Statistical significance levels: *$p < 0.01$; **$p < 0.001$.

Table A3: Mean covariate balance obtained from coarsened exact matching

	Treated	Control	Difference	Improvement
Distance	0.48	0.46	0.02	94%
Violent crime rate	229	210	19	88%
Median household income ($)	43,393	43,429	−36	99%
High school graduation rate	0.73	0.73	0	99%
Fraction African Americans	0.09	0.09	0.003	97%
Fraction Hispanics	0.03	0.03	0.003	92%
Fraction other ethnicity	0.01	0.01	0.0007	49%

Notes: Results from applying a coarsened exact matching algorithm to time-averaged versions of the variables presented in the Data and methods section. The algorithm produces a matched data set using a dichotomous treatment indicator of above versus below mean exposure to incarceration. The table displays matched covariate levels in the control and treatment groups, as well as the percentage balance improvement as a result of matching.

Table A4: Cross-sectional regression model of communicable disease mortality

Control variable	Control coefficient	Incarceration coefficient
Unemployment rate	4.0**	1.6**
	(0.3)	(0.3)
Poverty rate	5.0**	1.4**
	(0.2)	(0.2)
Absolute income mobility	−3.3**	1.5**
	(0.3)	(0.3)
Income inequality	0.9**	1.9**
	(0.3)	(0.3)
Racial segregation	0.8**	1.9**
	(0.2)	(0.3)
Percentage uninsured	1.3**	1.9**
	(0.3)	(0.3)

Notes: The outcome variable is mortality from communicable diseases in the year 2014. Each row is a separate regression wherein the association between incarceration and mortality is adjusted for the control variable listed in the first column. All models are also adjusted for state-fixed effects. All regressors are standardised by subtracting the mean and dividing by the standard deviation. Parameter estimates are interpreted as the excess mortality rate associated with a standard deviation increase in each predictor. Robust standard errors are shown in parentheses below each parameter estimate. Statistical significance levels: *$p < 0.01$; **$p < 0.001$.

Table A5: Cross-sectional regression model of non-communicable disease mortality

Control variable	Control coefficient	Incarceration coefficient
Unemployment rate	43.9**	34.9**
	(3.8)	(2.5)
Poverty rate	42.1**	34.7**
	(3.2)	(2.5)
Absolute income mobility	−23.7**	36.5**
	(4.0)	(2.8)
Income inequality	−3.2	39.8**
	(2.6)	(2.7)
Racial segregation	1.2	39.4**
	(2.0)	(2.7)
Percentage uninsured	5.2	39.3**
	(3.9)	(2.7)

Notes: The outcome variable is mortality from non-communicable diseases in the year 2014. Each row is a separate regression wherein the association between incarceration and mortality is adjusted for the control variable listed in the first column. All models are also adjusted for state-fixed effects. All regressors are standardised by subtracting the mean and dividing by the standard deviation. Parameter estimates are interpreted as the excess mortality rate associated with a standard deviation increase in each predictor. Robust standard errors are shown in parentheses below each parameter estimate. Statistical significance levels: *$p < 0.01$; **$p < 0.001$.

Table A6: Cross-sectional regression model of all-cause mortality

Control variable	Control coefficient	Incarceration coefficient
Unemployment rate	58.5**	39.0**
	(4.7)	(3.0)
Poverty rate	55.7**	38.8**
	(3.9)	(3.0)
Absolute income mobility	−27.4**	41.8**
	(3.5)	(3.4)
Income inequality	−6.5	45.7**
	(3.1)	(3.3)
Racial segregation	−0.6	45.2**
	(2.5)	(3.2)
Percentage uninsured	18.0**	44.4**
	(4.8)	(3.2)

Notes: The outcome variable is mortality from all causes in the year 2014. Each row is a separate regression wherein the association between incarceration and mortality is adjusted for the control variable listed in the first column. All models are also adjusted for state-fixed effects. All regressors are standardised by subtracting the mean and dividing by the standard deviation. Parameter estimates are interpreted as the excess mortality rate associated with a standard deviation increase in each predictor. Robust standard errors are shown in parentheses below each parameter estimate. Statistical significance levels: $^{*}p < 0.01$; $^{**}p < 0.001$.

Figure A1: Observed and imputed values of annual aggregate per capita correctional expenditure.