The Journal of Economic Perspectives

A journal of the
American Economic Association

Fall 2021

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@aeapubs.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by LSC Communications, Owensville, Missouri, 65066, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Fall 2021, Vol. 35, No. 4. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $24.00, $34.00, or $44.00 depending on income; for an additional $15.00, you can receive this journal in print. The journal is freely available online. For details and further information on the AEA go to https://www.aeaweb.org/.
Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 35 • Number 4 • Fall 2021

Symposia

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 3–28

The Economics of Policing and Public Safety

Emily Owens and Bocar Ba

Police are the first point of contact that most people have with the criminal justice system. In the interest of public safety, police in the United States have a broad mandate to define exactly what actions may constitute a crime and what behaviors suggest that someone may have participated in one of those actions. Governments grant police the right to engage with the public, which we define broadly as any police action that is intended to increase safety, including the right to display, threaten, or use force in order to coerce civilians into complying with police demands. These state-granted rights of engagement have both benefits and costs to society.

There is consistent evidence of a substantively large and negative elasticity of crime with respect to the number of police in an area (Levitt 1997, 2002; McCrary 2002; Di Tella and Schargrodsky 2004; Evans and Owens 2007; Draca, Machin, and Witt 2011; Fu and Wolpin 2018; Mello 2019; Weisburst 2018). There is also evidence that certain police practices can reduce crime, including problem-solving and (some) proactive policing strategies (National Academies of Sciences, Engineering, and Medicine 2018). Reducing crime is good for society. The direct costs associated with being a crime victim include lost wages, medical bills, and lost or destroyed property (for a recent review, see Bindler, Ketel, and Hjalmarsson 2020). In addition to direct costs of being a crime victim, civilians who are worried about the possibility of being victimized also incur costs of private actions taken to reduce their exposure to crime and costs associated with the stress of potential victimization

■ *Emily Owens is Professor of Criminology, Law, and Society, and Economics, University of California, Irvine, Irvine, California. Bocar A. Ba is Assistant Professor of Economics, Duke University, Durham, North Carolina. Their email addresses are egowens@uci.edu and bocar.ba@duke.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.3.

(Cohen 2014). While these indirect costs may be small in per capita terms, when summed across individuals, total indirect social costs are almost certainly larger than the total direct costs to the 1.68 percent of people who are crime victims each year (Bentham 1789).[1] Comparing estimates of direct costs to victims to survey evidence on people's willingness-to-pay to reduce their probability of victimization by, for example, living close to a registered sex offender, confirms this belief (Cohen 2014).

The ways in which police engage with the public to reduce crime include being present in a neighborhood, formal surveillance, issuing a citation, making an arrest, or using physical force to obtain compliance. Being the subject of police engagement is costly, in particular by decreasing the physical and mental well-being of the involved civilian (Geller et al. 2014; Geller and Fagan 2019; Mello 2018; Legewie and Fagan 2019; Harris, Ash, and Fagan 2020; Ang 2020). Other costs of police engagement are borne indirectly by others. An important component of these indirect costs include decreased trust in police, particularly if witnesses or communities may consider the engagement to be evidence of underlying illegitimacy of the police department or bias on the part of individual officers. As police officers are representatives of the government, decreased police legitimacy may also threaten civilians' trust and engagement with other political processes and governmental institutions (Lerman and Weaver 2014; Brayne 2014; Ba 2018; Ang and Tebes 2018). Much like the benefits of crime reduction, the roughly 24 percent civilians who are directly involved with the police each year bear the largest per capita costs.[2] We are not aware of empirical research on the indirect costs of police engagement—for example, how much do members of the public, on average, change their behavior in order to avoid police who may be racially biased? But the increased visibility of incidents of police aggression via cellphone technology and social media likely means these indirect costs are increasing (Owens 2019).

Some police actions, in certain contexts, will provide more benefit to society in terms of crime reduction than they cost in terms of police legitimacy, and some actions will not. Socially optimal policing occurs when police take the actions that provide a net social benefit, and when they refrain from actions that result in net costs. In practice, differentiating between costly and beneficial actions in the heat of an encounter between police and civilians is not easy, but describing optimal policing as merely balancing social benefits and costs actually understates the complexity of providing welfare-enhancing public safety through police.

This is because the distribution across people of the *direct* net benefits of police engagement is not obviously correlated with the distribution of *indirect* net benefits. Criminologists have long noted that fear of crime (an indirect benefit) is only weakly correlated with actual crime incidence (a direct benefit), and many police policies reduce either one or the other (Weisburd and Eck 2004). This pattern complicates the "ideal" provision of public safety through police; when the indirect benefits of

[1] This victimization rate estimated from the 2018 National Crime Victimization Survey (data series NCJ 253043).

[2] This estimate comes from the rate at which people in the 2018 Police Public Contact Survey (data series NCJ 255730) report being stopped by the police.

crime reduction are so large relative to the direct benefits, even a police action that is technically socially optimal can actually result in a welfare loss to people who are directly affected, as offenders and/or victims, by police engagement.

Policing does not exist in a vacuum, and the concentration of victims and offenders in US society is almost certainly the result of historic and persistent institutional discrimination. To the extent that police engagement occurs within communities already impacted by discrimination in other sectors, the marginal legitimacy cost of police engagement can be particularly high relative to engagement with groups of people who more firmly trust in the police (Hinton 2016). Police chiefs sometimes refer to the costs of using excessive force as "spending down community capital," a phrase which intuitively reflects this underlying idea. Not only is it the case that policies which appeal to a majority of voters and actually balance overall costs and benefits can be misprovided at local levels, the same people and places that bear the largest burden of misprovision of police engagement also tend to be the most historically disadvantaged. Attending to the distribution of direct and indirect net benefits adds an additional layer of complexity to the standard social welfare maximization problem.

As we will show, civilian feedback suggests that in practice, police engagement frequently is perceived by the public to reduce crime by less than it reduces legitimacy, and there is also evidence that police actions that provide net local benefits do not always occur. The local misprovision of policing is more likely to be reported by people who are Black or Hispanic. This outcome is not surprising when viewed in the context of the organizational and individual incentives of police.

External pressure on police departments tends to encourage crime reduction, with less attention to legitimacy costs. Even if a department was able to identify socially optimal police policies that carefully consider the distribution of direct and indirect benefits, it is not obvious those policies would be implemented by individual police officers. Current tools available to police managers to encourage individual officers to behave in the interest of the department are, from a personnel perspective, limited, and have been only rarely shown to alter police behavior in the field. Instead, many compensation and oversight strategies tend to encourage individual officers to make arrests and to emphasize officer's personal safety. While these goals are both important and laudable, they are different from engaging with the public in a socially optimal way.

Municipal Police Departments Are Funded like First Responders, Not Crime Reducers

Police departments are one of many government institutions, along with schools, social welfare organizations, and housing providers, that can benefit society by reducing crime. One way to think about how much municipal governments have chosen to provide public safety by increasing the cost of crime, versus reducing the perceived benefits of crime, is by comparing how many police there are relative to other local government employees, and how much those employees are paid.

In Figure 1, we compare the number of people employed by local police departments, primary and secondary schools, housing authorities, and social welfare organizations, from 1993 to 2016, as recorded in the Annual Survey of Public Employment Payroll. We also include the employment of firefighters: as first responders in emergencies, police and firefighters are tightly linked in the public mind. From 1993 to 2016, the numbers of people employed by police departments, schools, and fire departments have grown at roughly the same rate—29 percent for police and educators and 35 percent for firefighters. The number of people employed in social safety net organizations that have also been shown to reduce crime has not kept pace. Overall, in terms of number of people employed, it does not appear to be the case that municipal governments have reduced their investment in education relative to deterrence over time.

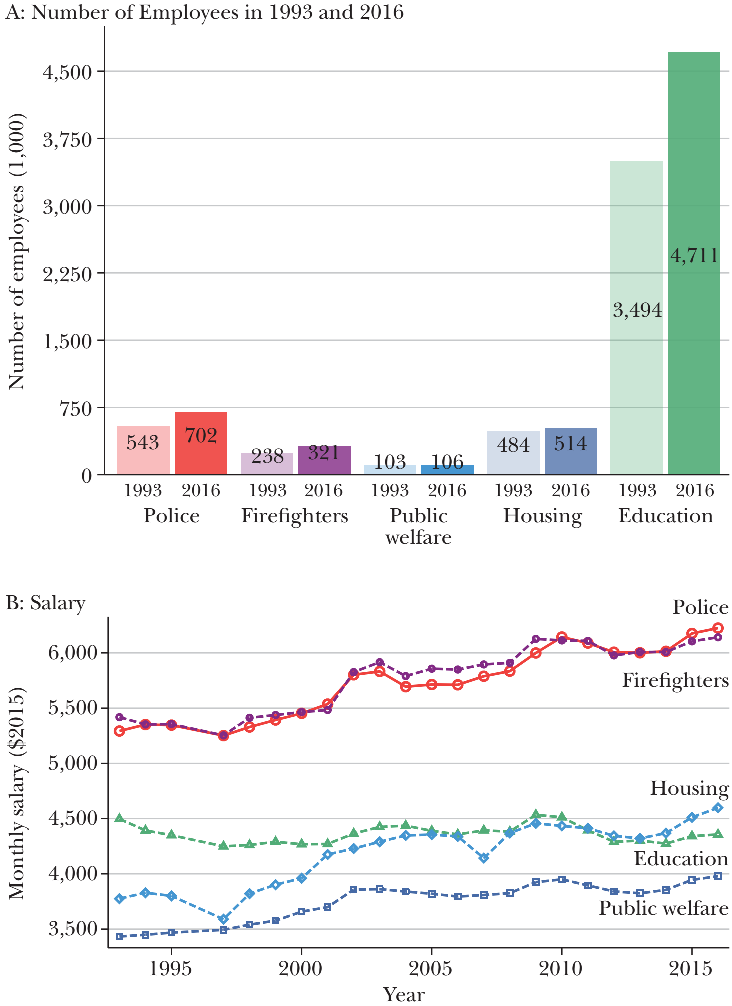
In contrast, if local investment is inferred from growth in salary, police officers have been treated more like firefighters than other public employees. Monthly police salaries have grown in lockstep with firefighters since 1993. In contrast, educators' salaries have been stagnant, or even falling. Trends in wages over time suggest the budgeting decisions regarding paying and hiring police assume their primary role in the community is to respond to emergencies. However, other first responders do not wield the discretionary influence that police officers have in the daily lives of civilians. From a budgetary perspective, allocating funds to officers in a way that focuses only on their role as first responders oversimplifies their role in society—as noted earlier, rapid response to crimes that have already occurred is only one of the many benefits of police engagement (Weisburd 2018).

What Do Civilians Think about Police?

One way to evaluate the extent to which policing decisions are maximizing social welfare is to ask people about their perceptions of law enforcement. According to Gallup polling, between 1995 and 2013, the percent of adults with a great deal of confidence in the police was between 54 percent and 64 percent. Since 2013, this has fallen, and in June 2020, only 48 percent of adults expressed strong support for law enforcement. In the 1990s and early 2000s, there was a persistent racial gap in public confidence, with about 60 percent of White Americans and 35 percent of Black Americans expressing a great deal of confidence in the police. In 2020, this racial gap grew (primarily due to reduced confidence among Black Americans) to almost 40 percentage points (as reported by Jones 2020).

The persistent and increasing difference in the amount of confidence that Black and White Americans have in police is not consistent with crime reductions being the only dimension along which police impact people's lives. Since 1993, crime in the United States has declined enormously. Sharkey (2018) points out that victimization was once a central fact of life for people living in large cities, and this is simply no longer the case. In addition, Black people have disproportionately benefited from the great crime decline, with violent victimization rates falling from 6 percent for Black adults in 1995 to 1.9 percent in 2019. For White adults,

Figure 1
Public Employment from 1993 to 2016



Source: Annual Survey of Public Employment Payroll (Kaplan 2021).
Note: These panels provide public employment patterns from 1993 to 2016. Data are missing for the year 1996. Panel A presents the number of full-time employees in 1993 and 2016. Panel B shows salary trends by sectors for full-time workers. For police and fire protection, we report the numbers associated with officers and firefighters. Elementary and secondary instructional comprise the education sector. The housing sector includes community development.

the comparable change was 4.3 percent to 2.1 percent. There are many contributors to the great crime decline of the 1990s, but an increase in police engagement, as measured by the number of police per capita in the United States, was almost certainly an important one (Levitt 2004). However, the 60 percent reduction in crime victimization risk was not enough to increase the confidence that Black adults had in police officers. As economists, this should not be a total surprise, because simply quantifying the crime-reducing benefits of policing informs only half of the crime-legitimacy tradeoff that police officers must make. Indeed, Black Americans are also more likely to bear large direct costs of police engagement: based on the Uniform Crime Reporting data over the same time period, adult arrest rates were at least twice as high for Black Americans than White Americans, and have only slightly converged.[3]

In order to better understand the temporal and racial dynamics of trust in the police over such a long period of reduced victimization risk, we now provide descriptive information about allegations of misconduct by officers from the Philadelphia Police Department (PPD) between January 2015 and November 2019 (for details about the complaint process, see PAC 2021). Philadelphia is the sixth-most populous city in the United States, but here, we focus on Philadelphia for reasons of data availability: we can observe the race of the complainant and what they specifically were complaining about. Filed complaints are an imperfect measure of officer performance, as we will discuss, but for the purposes of this section we interpret the act of filing a complaint as an expression of a preference for a different sort of police engagement, even if supervisors ultimately decided there was no evidence that the officer violated departmental policy.

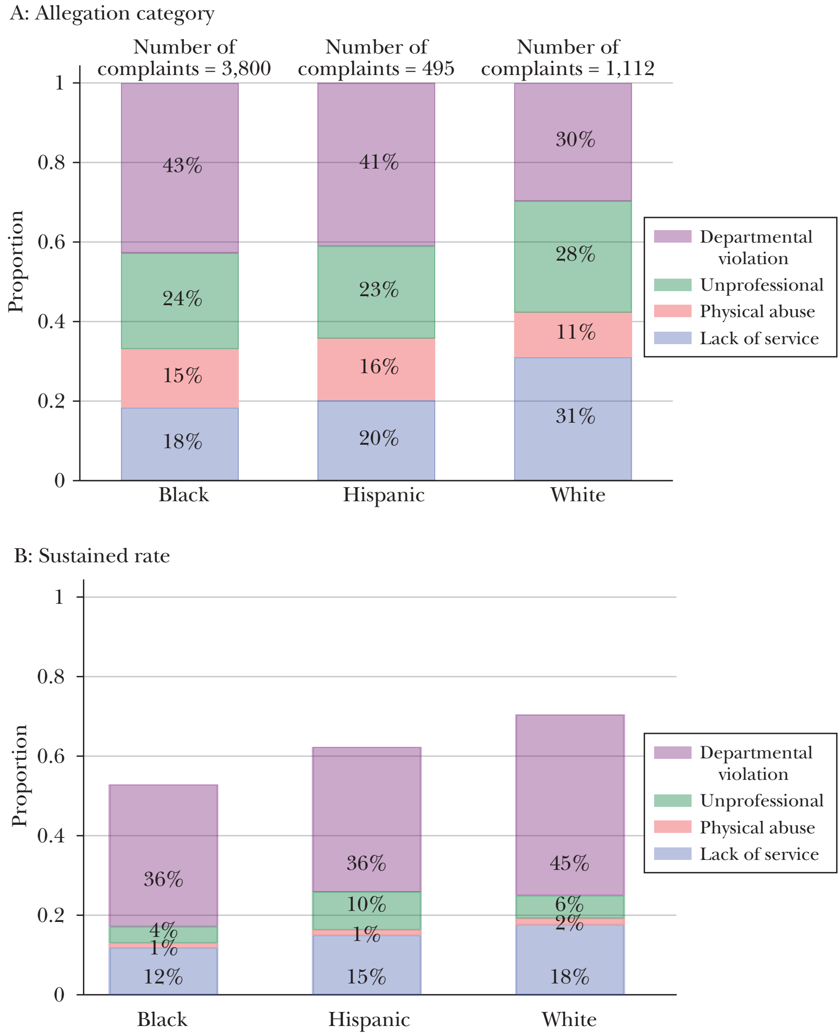
Among the 6,300 sworn officers in the Philadelphia Police Department, about 13–16 percent receive at least one complaint each year. According to Figure 2, civilians most frequently complain about police engagement that provides too little crime reduction relative to the imposed costs, 36.7 percent of the complaints allege a departmental violation or policing outside of the law, and 25 percent of the complaints allege verbal and/or physical abuse, all of which correspond to over-policing. However, concluding that the Philadelphia Police Department should simply reduce the amount it engages with the public is not clear: 20.5 percent of the complaints allege a lack of service, meaning that the civilian believes that not enough engagement was provided and would be a sign of under-policing).

Relative to the population of Philadelphia, Black people are overrepresented among complainants. About 60 percent of the complaints were filed by Black complainants, who make up 40 percent of the city population, and about 17 percent of the allegations were filed by White complainants, who make up 34 percent of the population. Taken at face value, Black Philadelphia's residents appear to be less satisfied with the quality of policing provided to them.

Whether civilians perceive over- or under-policing as the main problem varies by race. Almost 40 percent of Black civilian complaints allege a departmental violation while only 17 percent allege a lack of service. This pattern of complaints made by Hispanic civilians is similar, but inverted for White complainants. Almost one-third (27 percent) of complaints made by White civilians allege a departmental violation while 30 percent allege an officer's lack of service. When it comes to verbal and/or physical abuse, Black and Hispanic complaints make a higher

[3] OJJDP Statistical Briefing Book. Trends in arrest rates by race for all offenses (rates are per 100,000 in age group). Available: https://www.ojjdp.gov/ojstatbb/crime/ucr_trend.asp?table_in=2. Released on November 16, 2020.

Figure 2
Complaint Patterns by Complainant's Race in Philadelphia



Source: Open Data Philly
Note: These figures provide information on allegations of misconduct by Philadelphia police officers between January 2015 and November 2019. We restrict the analysis to Black, Hispanic, and White complainants. Panel A presents the distribution of the allegation category. Panel B presents the sustained rates for investigated allegations, specifically, complaints that might lead to potential disciplinary action or training.

share of allegations (about 26–27 percent) compared to White complainants (23 percent). Overall, Black and Hispanic civilians are more likely to perceive that officers are overpolicing—taking relatively more actions that do not reduce crime by as much as they erode trust in the police. In contrast, White civilians are more likely to demand that police officers engage in more crime-reducing actions.

Are the complaints indicative of rogue officers or dissatisfaction with department policies? One way to understand this is to consider what happens after a complaint is made. Officers who have complaints filed against them are not immediately sanctioned; instead, complaints generally trigger an investigation process where the officer's supervisors determine if the officer violated policy. Of the overall complaints in our sample, only 18.1 percent are sustained; over 80 percent of the time that a civilian complains, the officer's actions were considered to be consistent with how their supervisor expected them to behave.[4] The fact that officers are determined to be out of policy in fewer than one out of five complaints filed is consistent with a misallocation of police engagement at the organizational level, rather than individual officers not following departmental guidelines. Officers appear to be acting in accordance with what their superiors in the department have instructed them to do, but those instructions do not appear to be delivering the balance of crime reduction and civil rights protections that the public desires. This appears to be of particular concern for the policing of Black civilians, whose complaints are sustained the least often.

This case study of civilian complaints in Philadelphia suggests that Black and White civilians have different preferences about current police engagement. However, it is not obvious that one group wants more engagement while another wants less; White civilians are more likely to demand more police engagement (complaining about lack of service at three times the rate as physical abuse), and Black and Hispanic civilians appear to primarily demand different type of engagement than what is being offered, complaining about under- and over-policing at roughly equal rates. In the remainder of the paper, we describe how this misprovision of services, which varies across racial and ethnic identity, can be understood as an outcome of the structure of incentives facing police departments and individual officers.

What Are the Organizational Priorities of the Police?

In practice, the elected officials who lead or appoint the leaders of policing agencies have two incentives: 1) to provide the type of police engagement preferred by the median voter; and 2) to raise municipal revenue. Neither of these goals is necessarily consistent with striking the appropriate balance between under- and over-policing, or with attention to the varying incidence of the indirect and direct net benefits of police engagement. The organizational priorities of local police can also be affected by incentives created by the federal government and the communities they serve. The federal government can manipulate the choices of local policing agencies, and primarily does so through providing grants that subsidize specific police practices or conditioning access to federal resources on certain types of community engagement. Community preferences can also lead departments to adopt different kinds of police engagement; qualitative evidence suggests that at

[4] The categories of complaints sustained most often are departmental violations and officer lack of service. The category that is sustained the least is for physical abuse (less than 2 percent).

least some of the mechanisms through which departments solicit actual community feedback is more likely to include community members who favor more police engagement in an absolute sense.

Elected Officials and Re-election Incentives

Policing in the United States is highly decentralized; the primary law enforcement entity in most of the United States is a municipal authority. Local elected officials are responsible to voters.

A voter's beliefs about the cost of crime likely differ from their objective risk of direct victimization. For example, while US crime rates were falling dramatically after 1995, a majority of Americans have consistently reported thinking that the crime problem in the United States was getting worse each year (Sharkey 2018). Hiring more police officers is a salient example of the local government "doing something" about crime (Webb and Katz 2014), so much so that Levitt (1997) was able to use the timing of mayoral elections as an instrument for the number of police in an early paper identifying the causal impact of police staffing on crime. But do these electoral incentives improve the social efficiency of policing? Ornaghi (2019) finds that limiting the ability of mayors to influence police personnel decisions leads to substantial reductions in reported property offenses. This suggests that political pressure on law enforcement may be focused on providing indirect benefits—taking a visible action that increases perceived safety—a goal which may only overlap in part with actually providing direct benefits of reduced crime.

To the extent that race provides a rough proxy for the incidence of direct costs and benefits of police engagement, the aftermath of the 1964 Civil Rights Act provides an example of how changes in the electoral power of people directly affected by policing influences departmental incentives. The Civil Rights Act increased the ability of Black Americans to vote in specific counties, making the preferences of Black civilians about policing more important in local elections. Facchini, Knight, and Testa (2020) finds that (elected) sheriff's deputies reduced their level of engagement in response to this increased electoral power. Specifically, in the affected counties, fewer Black people were arrested for low-level offenses—the kinds of incidents for which the benefit of crime reduction is likely lower relative to the direct legitimacy cost of arrests. When political power is concentrated among people who are not directly affected by crime and police, elected officials do not have an incentive to identify or promote police actions that maximize the net direct benefits of police engagement.

Police Engagement Affects Local Municipal Budgets

Issuing citations or making arrests for offenses that are penalized by fines can provide revenue for local government officials. Criminal fines generally accrue to the local government through the court system. Anecdotal concerns about police issuing traffic tickets as a means of revenue generation are increasingly supported by some empirical evidence (Makowsky and Stratmann 2009, 2011; Makowsky, Stratmann, and Tabarrok 2019). One notable finding of this emergent literature is that municipal budget deficits seem to affect policing enough to motivate very specific types of engagement. Traffic citations, arrests for sex work, and drug arrests of Black and Hispanic people appear to be the most responsive to local budget needs. An extension of this result, if confirmed by later research, is that the individual police officers who issue these citations and make these arrests do not always view their safety-enhancing benefit as sufficient to outweigh the costs—instead, the municipal financial need is pivotal.

Of course, police engagement can also create costs for the government, which we will discuss in detail in the next section. Most cost-benefit calculations of policing in the economics literature focus on salary and pension obligations but do not account for monetary costs associated with police behavior that could impact credit risk for local governments (for example, Chalfin and McCrary 2018). A recent report by Moody's noted that police operations contribute several credit risks, including high police pension burdens, expensive settlements and consent decrees, and expenditure pressures (Strungis et al. 2020). In addition, the extent of police misconduct also affects taxpayer-funded settlements and reform efforts (Schwartz 2016; Rushin 2017; Ouss and Rappaport 2020). Moody's also notes that poor quality of public safety is a signal of city defaults and bankruptcy, as in the case of Detroit, Michigan, and Stockton, California (Strungis et al. 2020). There is some evidence that municipalities may encourage police engagement in order to raise revenue, but it is unclear how much local governments constrain police actions in order to minimize budgetary costs.

Departmental Incentives Created by the Federal Government

The federal government can and does affect the decisions made by local law enforcement. There are many reasons why the federal government might want to do so. Policies adopted by individual municipal police departments can have implications for the well-being of residents outside of its jurisdiction—this is particularly true for cities with far-reaching suburbs. To the extent that high crime rates are one consequence of an underfunded municipality, the federal government may subsidize certain law enforcement decisions (like hiring) that more fiscally constrained local governments are unable to afford. The US Constitution also places legal constraints on how police officers, as government agents, can interact with civilians, most notably through the Fourth (regulating police stops and searches) and Fourteenth Amendments (requiring equal protection of law).

Of course, there are also reasons to limit the federal government's ability to manipulate decisions about local policing. Local governments tend to be more responsive to heterogeneity in the demand for police engagement, for example, where there is geographic variation in preferences about the use of intoxicating substances. Even further, experimental evidence from criminology suggests that understanding the specific local causes of a crime problem is critical to solving it. The strategies whose crime-reducing benefits are most generalizable to different contexts are the least prescriptive: as one example, the Scanning, Analysis, Response, and Assessment model (SARA) directs officers to identify what the particular cause of conflict is in a specific place and subsequently take steps to address that issue (National Academies of Sciences, Engineering, and Medicine 2018).

In practice, the primary way in which the federal government affects local policing is by creating financial incentives for departments to enact specific policies. Historians have documented the emergence of federal control of local law enforcement through grant-making in the Johnson and Nixon administrations, shifting the local response to crime from investments in education and job creation to increased police engagement, particularly in Black neighborhoods (Hinton 2016). In addition, the high rate of voluntary participation with the Federal Bureau of Investigation's Uniform Crime Reporting program can be attributed, at least in part, to the fact that these data are used to calculate eligibility for federal block grants that go towards policing (a program which has existed under various names since the 1980s).

Most economic research on federal grants for law enforcement has focused on the role of the Violent Crime Control Act of 1994 in increasing the number of police officers (Evans and Owens 2007; Cook et al. 2017; Weisburst 2018; Mello 2019) and the number of police officers placed in schools (Owens 2017; Weisburst 2019). The law accomplished this by offering short-term wage subsidies for newly hired officers through the Community Oriented Police Services office. These subsidies were found to be associated with increases in the number of local police, and substantial reduction in crime, which is consistent with, but not necessarily proof of, previous underinvestment in police by local governments.

The federal government has more recently used financial incentives to increase local law enforcement's efforts on people who may be in violation of the 1952 Immigration and Nationality Act. Participation in Secure Communities, a federal program aimed at identifying and detaining immigrants held in local jails who lack legal authority to remain in the United States, is necessary for local authorities to be able to access the national fingerprint repository maintained by the Federal Bureau of Investigation. In addition, the federal government can condition the receipt of federal grants for policing, specifically, the Edward Byrne Memorial Justice Assistance Grant Program, on a local law enforcement's cooperation with Immigration and Customs Enforcement.

Involvement of local law enforcement with immigration enforcement is an area where the tension between local and federal incentives is perhaps the most salient; being able to take advantage of the specific knowledge that local law enforcement has about the community clearly benefits federal immigration officials who seek to identify and remove people in violation of federal immigration law. However, aggressive enforcement of immigration law may reduce the willingness of immigrants, particularly those without legal authority to remain in the United States, to cooperate with local police in preventing local crime (Comino, Mastobuoni, and Nicolò 2020; Jácome 2020).

Departmental Incentives Created by Community Members

Police were discouraged from regularly engaging with the public in non-enforcement contexts under the "professional" model of policing, pushed in the first half of the 20th century by leaders like August Vollmer and O.W. Wilson. One of the responses to the 1980s crime wave was a return to "community oriented policing," in which departments seek input from the people they serve in identifying the optimal amount and form of police engagement. Examples of community oriented policing activities include: 1) holding regular meetings with community members to discuss crime-related issues, generate potential solutions to those problems, and evaluate the efficacy of those solutions (commonly referred to as the SARA model); 2) forming a partnership with a local business or neighborhood organization; 3) making officers available to the public at specific times and locations; or 4) including local residents in policing activities by holding "citizens academies" or civilian-police neighborhood safety walks. According to the 2016 Law Enforcement Management and Administrative Statistics (LEMAS), 42 percent of all US law enforcement agencies, and more than 70 percent of the agencies serving populations of more than 250,000 people, have written policy statements that explicitly incorporate community-oriented policing.

However, the information gathered from community-oriented policing may not be socially optimal from a local or jurisdictional perspective. As ethnographers have documented, not all members of a community are equally likely to participate in outreach events. Individuals who show up for community meetings are more likely to be White and wealthier than the average resident, and members of marginalized groups who have had negative encounters with police officers can feel particularly unwelcome in these spaces (Muniz 2012). Moreover, officers in these meetings may be more likely to formally document community requests for increased police engagement than requests for lower levels of engagement (Cheng 2019).[5]

What Causes Officers to Behave in the Department's Interest?

Supervising the individual police officers who interact with the public is a complicated and high stakes personnel problem. Individual police officers are people with their own set of incentives, which may differ from the objectives of their employers. To the extent that people who choose to become police officers are interested in promoting the well-being of society, one would expect officers to think carefully about the legitimacy costs and crime-reducing benefits of their engagement with the public, perhaps asking themselves, "what would the people I serve like me to do right now?" However, potential police officers who were attracted to the job because of their desire to be forceful warriors may provide

[5] This selection issue is even more prominent when policing is provided by, or in cooperation with, the private sector. Security guards employed by private companies or associations can frequently provide more targeted and "nimble" engagement than would be possible for a governmental agency, and can sometimes provide more crime reduction than the existing public police (Brooks 2008; Cook and MacDonald 2011; Heaton et al. 2016; Cheng and Long 2018). This increased crime reduction is consistent with private interests bearing relatively less of the social legitimacy costs of police engagement, as a government should. In addition, private officers are not always bound by the same constitutional restrictions that police are, in particular Fourth Amendment limits on search and seizures (Meares and Owens 2019).

more engagement than a benevolent social planner would prefer. Currently, the standard set of tools available to police departments to solve this principal-agent problem is limited. As of 2021, the existing evidence base on how different screening, training, and monitoring strategies affect officer behavior in the field is minimal, let alone research on the specific question of what personnel strategies help officers to better balance legitimacy costs with crime-reducing benefits. That said, we can explore, using the 2016 Law Enforcement Management and Administrative Statistics Survey, what personnel tools departments currently use. We focus on departments that employ more than 100 officers, with jurisdictions that are large enough to have demographic information reported in the 2018 American Community Survey.

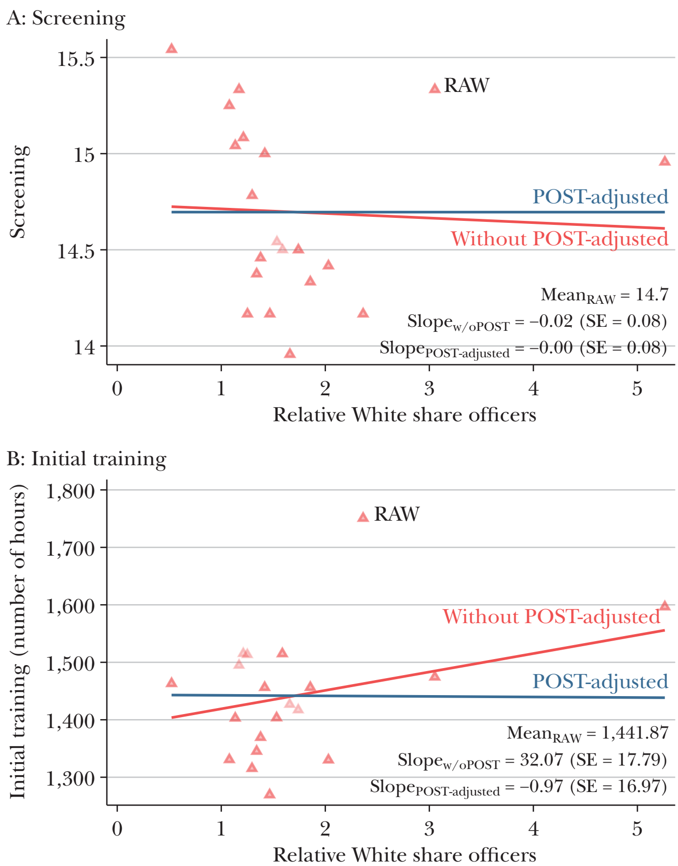
Screening Job Candidates

On average, the police departments in our sample use almost 15 tests to screen candidates. Many of these screening practices are recommended or required by a state-level oversight body, commonly a Peace Officer Standards and Training (POST) commission. Requiring officers to have attended some college is rare, but multidimensional screening is common. Of the 485 police agencies for which we have city-level demographic information, 27 percent require more than a high school degree, and the departments on average use 4.8 (out of 5) different types of background checks, 5.17 (out of 7) personality tests, 0.6 (out of 2) tests of ability to manage community relationships, and 3.86 (out of 4) physical tests. Very little is currently known about whether these tests identify the type of officers desired by the public.

One additional screening tool available to departments that, while not included in the Law Enforcement Management and Administrative Statistics Survey but remains the subject of policy attention, is the use of racial and ethnic identity in the selection of officers. Diversifying the police force is often proposed as a solution to improve police-civilian interactions (Ba et al. 2021b), but its impact is difficult to assess and is sometimes criticized as requiring a reduction in other important hiring standards. In contrast, Figure 3 suggests police departments that use more of the screening procedures measured in the Law Enforcement Management and Administrative Statistics Survey do not appear to have "Whiter" police forces, which is not consistent with the assertion that efforts to increase police officer diversity necessarily require a relaxation of other standards.

Unlike other screening tools, there is some evidence that more diverse departments are able to police with lower legitimacy costs, and weakly lower crime rates. Donohue and Levitt (2001) document that same-race policing is associated with a reduction in the number of arrests where the officer and the suspect appear to be of the same race, especially for minor offenses. Several studies have considered the after effects of court-ordered affirmative action programs, which can be viewed as an exogenous shock to the share of officers who are Black and/or female. McCrary (2007) does not find evidence that crime rates were affected by successful affirmative action lawsuits, but arrests per crime and the number of Black civilians among people arrested for serious crimes dropped. Further, Harvey and Mattia (2019)

Figure 3
Screening and Training



Source: 2016 Law Enforcement Management and Administrative Statistics.
Note: These panels show the relationship between officers' training and screening as a function of the relative White share officers. Each binned scatter plot (RAW) is constructed by dividing the data into 20 equal- sized bins, ranking by the *x*-axis variable. The relative White share officers is the ratio between the share of White officers in the jurisdiction divided by the share of White residents. Each panel reports the slope coefficient and standard error of the corresponding linear ordinary least squares regression with and without controlling for the Peace Officer Standards and Training (POST) dummy. For regression details, see online Appendix available with this article at the *JEP* website.

show Black crime victimization decreased after the previously mentioned court orders. Miller and Segal (2018) also find that the number of crime reports involving violence against women increases, reducing the actual rates of domestic violence, after lawsuits intended to increase the employment of female officers.

While diversifying policing does seem to make a difference at the city-level, less is known about the micro-level. The tasks given to police officers vary on a day-to-day basis, often according to gender and race. For instance, Ba et al. (2021b) document that Black officers in Chicago work different shifts, districts, and beats than non-Black officers. They also find similar patterns across different genders. As a result, actions taken by police of different races or genders may reflect both variation in who officers in different identity groups interact with, as well as how those different officers make engagement decisions. When comparing officers who work in similar shifts, districts, and beats, the authors find that minority officers make far fewer stops and arrests and they use less force relative to their White counterparts, especially when interacting with Black civilians. There is no evidence that Hispanic officers exhibit different behaviors when interacting with co-ethnic civilians relative to their White counterparts.

Training Cadets

Conditional on passing the department's screening tests, officers must complete an average of 1,441 hours of training, in both academy setting and in the field. Training is heavily influenced by the state-level Peace Officer Standards and Training commissions; in 2013, 95 percent of training academies were certified by POST or other state certification agencies, and 93 percent of agencies use POST developed curriculum for basic training. POST certification can also be awarded directly to trainers: 73 percent of training academies require full time trainers to be POST-certified (27 percent of the academies who are not certified on their own require full time trainers to be POST-certified).[6] Rather than being substitutes, each additional screening measure used by a department is associated with a 2.3 percent increase in required hours of training, which appears to be driven entirely by variation in state standards; within a state, there is essentially no correlation between the number of screening tests used and the number of training hours required. Evaluation of the quality of academy courses is generally based on cadet performance on an in-class exam, or virtual simulation environment, rather than the on-the-job performance of officers who completed the course.

Unlike screening, we find some evidence that departments with more mandated training by their Peace Officer Standards and Training commissions have departments that are disproportionately White; without accounting for state-level POST standards, 50 more hours of training is associated with a 0.6 percent increase in the relative Whiteness of the police force.[7] To the extent that racial composition matters, this may place downward pressure on the ability of those departments to reduce crime at low cost. However, we do not find evidence that departments that require more training than is mandated by POST are more relatively White than other departments in their state.

Monitoring Officers on the Job

Once screened and trained, police officers are tasked with implementing the policies and priorities set by their department's command staff, and that

[6] Author's calculations from the Census of State and Local Law Enforcement Training Academies, 2013 (Bureau of Justice Statistics 2018).

[7] The mean ratio of the percent of a police force that is White and the percent of the jurisdiction that is White is 1.711. On average, White Ratio = 1.40 + .0002122 x Initial Training, and (50 x .000212)/1.711 = .0058.

implementation is largely unsupervised in real time. In most departments, officers directly report to a sergeant who is generally in charge of multiple officers working a shift. After a sergeant's "roll call" meeting at the start of each shift, officers operate with very little direct oversight of their specific tasks. This leaves open the possibility of individual officer preferences playing a large role in their decisions. For example, to the extent that this discretion may include multiple forms of racial bias, it can distort that officer's community engagement away from the departmentally preferred or socially optimal balance of crime reduction and legitimacy (Lum et al. 2020; Knowles, Persico, and Todd 2001; Durlauf 2006; Goncalves and Mello 2017; Knox, Lowe, Mummolo 2020).

Officers who observe potential criminal activity are expected to intervene "on view," although it is not always obvious that command staff would be aware of failure to intervene in a situation that an officer discovered. Police officers also learn of potential crimes from dispatchers (who are called by civilians). Dispatchers affect how much discretion an officer has in responding to an incident. A low priority event, like a report of a small amount of stolen property, may be something where an officer is expected to be on scene as soon as is practical given other priorities. Top priority events, like a violent crime in progress, must be responded to immediately. Dispatched calls are, almost as a rule, observed by an officer's supervisor, and so the individual officer has less discretion over their level of engagement than in an "on view."

After a potential crime is reported to or observed by an officer, the officer is expected to identify and respond to any likely perpetrator(s), thus "clearing" the incident. Incidents can be cleared in many ways: determining there was no appropriate police response, rendering assistance, issuing a verbal warning or written citation, or making an arrest. The last form of clearance—clearance by arrest—tends to be rare. Both the official record of a crime and how it was cleared are the basic administrative record of an officer's daily activities. A combination of clearance rates and local crime reductions are frequently used as performance metrics by supervisors. Clearances by arrest (and only by arrest) are reported to the FBI as part of the Uniform Crime Reports, which in turn is covered in the news media.

Technological change has dramatically increased the ability of departments (and the public) to monitor police officers on the job, because more of what officers do is recorded electronically. On average, departments had 5.67 different electronic databases that measured different aspects of officer engagement in 2016, including crimes responded to, complaints, arrests, stops, motor vehicle accidents, and uses of force. While there are start-up and maintenance costs for this increased use of technology to record officer activity, it holds enormous potential in terms of minimizing principal-agent problems, which conversely are exacerbated by reliance on paper records.

One particular technological innovation in policing in the 1990s was the introduction of CompStat, initially created by the New York Police Department. CompStat is essentially a regular compilation and reporting of crimes and arrests by beat to police supervisors. Sherman (1998) found that prior to the introduction of CompStat, there was an average of 90 days between when an NYPD officer made an arrest and when the captain for that officer became aware of it. Digitization reduced that lag to roughly one week. The ability of command staff to track non-arrest actions is both more nascent, and arguably more important, than electronically tracking arrests, given that fewer than 10 percent of most police-civilian interactions end in an arrest (Owens et al. 2018).

In addition to monitoring capacity, the 2016 Law Enforcement Management and Administrative Statistics Survey also asks about the actual use of a particular type of officer supervision—an Early Intervention System, sometimes called an Early Warning System. An Early Intervention System monitors potentially problematic officer behavior that is captured in an agency's administrative data and can include both the earlier mentioned records of field activity as well as personnel information like overtime and approved secondary jobs. Officers who appear to be unusual in any one of these dimensions (or an aggregated value) relative to a predetermined peer group can be flagged for a supervisory meeting and potential department action. In 2016, over 70 percent of large police agencies used an Early Intervention System: since 1998, the US Department of Justice has recommended that agencies adopt one. Again, despite the increased use of an Early Intervention System and similar approaches, there is scant evidence, primarily consisting of case studies, evaluating whether adopting these systems reduces potentially problematic engagement by police officers (Shjarback and Nix 2020).

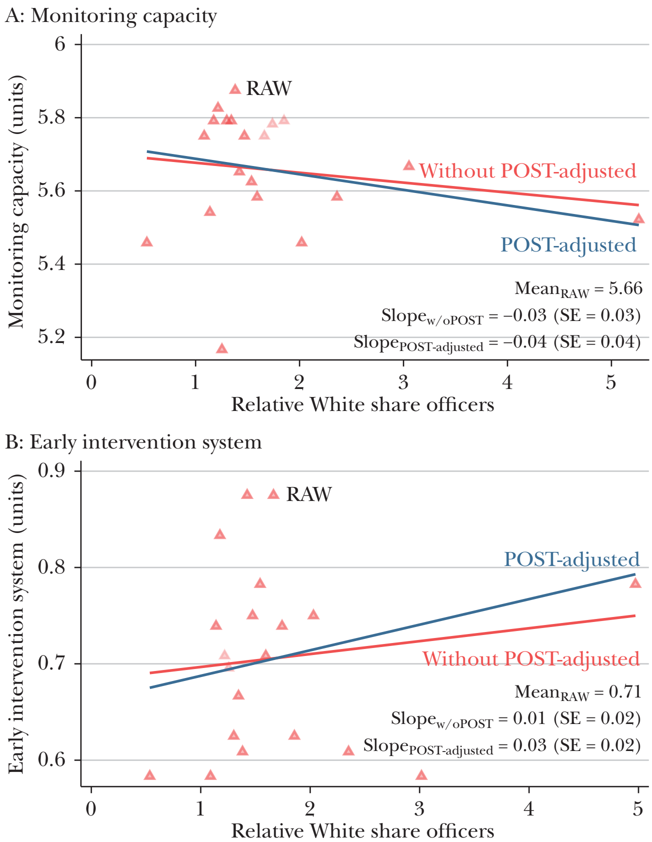
Given the potential role that officer race plays in both perceived and actual legitimacy costs and crime-reducing benefits of police encounters, command staff overseeing White officers working in non-White communities might disproportionately benefit from additional information about those officer's actions (Ba et al. 2021). Despite this potential benefit, as shown in Figure 4, we do not find evidence that departments where the cost of principal-agent problems may be larger currently have either higher capacity for electronic monitoring or are more likely to monitor their officers with an Early Intervention System in practice.

Civilian Feedback on Training and Monitoring

There is currently little credible causal evidence from the field on whether-screening, training, or monitoring strategies can provide officers with the tools they need to identify the socially optimal level of engagement in any particular civilian encounter. Until the research community provides such evidence, departments do have the option of soliciting feedback from the community about what sort of policies or practices they would like to see. In 2016, just over one-third of large agencies solicited civilian feedback about the performance of specific officers (35 percent), or potential officer training initiatives (24 percent), via a formal survey.

As shown in Figure 5, departments whose officers are less representative of the population appear to be less likely to seek out community input regarding training and monitoring. To the extent that the racial composition of these departments puts them at greater risk for allegations of bias or perceptions of illegitimacy, the observation that these same departments are also less likely to have a mechanism to identify civilian preferences on how officers are trained or supervised creates a large scope for socially inefficient policing to persist.

Figure 4
Monitoring Capacity and Early Intervention



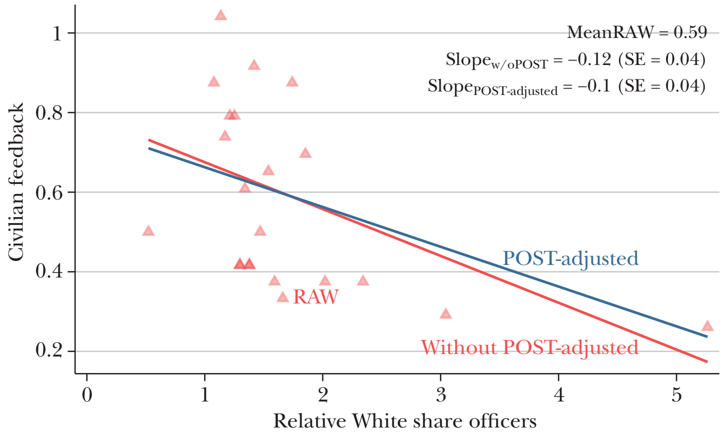
Source: 2016 Law Enforcement Management and Administrative Statistics.
Note: These figures show the relationship between monitoring capacity and early intervention as a function of the relative White share officers. Each binned scatter plot (RAW) is constructed by dividing the data into 20 equal-sized bins, ranking by the *x*-axis variable. The relative White share officers is the ratio between the share of White officers in the jurisdiction divided by the share of White residents. Each panel reports the slope coefficient and standard error of the corresponding linear ordinary least squares regression with and without controlling for the Peace Officer Standards and Training (POST) dummy. See online Appendix for regression details.

Payment and Promotion

A final mechanism departments could use to encourage officers to implement departmental policies is through pay, promotion, or sanctions. Unfortunately, policing institutions are generally limited in their ability to use these reactive mechanisms to alter officer incentives.

While there are certainly exceptions, pay and promotion in law enforcement agencies, like all government agencies, is standardized, and can be subject to union rules. Pay is generally based on tenure (Ba et al. 2021a) and, as noted in Figure 1, has tended to track other first responders rather than compensation or employment in other social welfare organizations. After a certain number of years in their

Figure 5
Civilian Feedback



Source: 2016 Law Enforcement Management and Administrative Statistics.
Note: This figure shows the relationship between civilian feedback as a function of the area demographics and the relative White share officers. Each binned scatter plot (RAW) is constructed by dividing the data into twenty equal-sized bins, ranking by the *x*-axis variable (that is, racial concentration and relative White are officer). The racial concentration is calculated by squaring the share of each racial group (Black, Hispanic, White, and other) for residents in the jurisdiction and then summing the resulting numbers. The relative White share officers is the ratio between the share of White officers in the jurisdiction divided by the share of White residents. Each panel reports the slope coefficient and standard error of the corresponding linear ordinary least squares regression with and without controlling for the Peace Officer Standards and Training (POST) dummy. For details of regression, see online Appendix.

current position, officers generally become eligible to take a test on law and police procedure. This test is sometimes written by the state-level Peace Officer Standards and Training commissions or an equivalent agency. There is typically a minimum qualifying score an officer must earn on this exam to become eligible for a promotion. When a position becomes available, promotion decisions are then based on at least one round of interviews with candidates who have met the testing threshold. It is at the interview stage that traditional performance metrics like clearance rates and arrests could come up along with the broader social welfare impacts of that officer's decision in the field.

As a general rule, supervisor feedback on specific tasks is "stick-based" and occurs after an officer is involved in a potentially problematic event: a complaint is filed or an officer uses force or engages in a vehicular pursuit. Finding that an officer violated policy can be grounds for dismissal, although only in rare cases are officers "decertified" by the Peace Officer Standards and Training commission—meaning that other agencies would be made aware of the reason for job separation. A few papers look at how officers respond to departmental sanctions (Prendergast 2001; Benoît and Dubra 2004; Rozema and Schanzenbach 2018), and in practice, the expected cost of a sanction, given that an officer violates a departmental policy, is probably quite low. This is particularly true when union regulations generally require a high burden of proof before sanctioning officers. Officers who are not involved in potentially sanctionable events will generally not receive specific feedback, although Owens et al. (2018) finds that non-investigative performance reviews can potentially reduce the rate at which officers choose to engage with an arrest or use of force.

As previously mentioned, the Fourth Amendment to the US Constitution defines how and when officers can restrict the liberty of the public. However, the direct cost of Fourth Amendment violations primarily fall on prosecutors who cannot use evidence gathered during an illegal search or seizure in court. A finding by federal investigators of a "Pattern or Practice" of repeat Fourteenth Amendment violations within a department can lead to a federal monitor asserting control over agency decisions. The increased oversight follows creates substantial time costs for command staff. The federal monitor can require that departments institute policies that impact officers, like requiring additional training, reporting requirements, or departmental reviews. However, the presence of a federal monitor and costs associated with implementing policies adopted during the consent decree likely impose a second order impact on the daily tasks of beat officers relative to the impact that the monitor has on their supervisors. Overall, while the Constitution does allow for federal oversight of individual officers, in practice the "sticks" it creates primarily do not affect individual officers. Further, in many jurisdictions police officers are formally protected from being held personally liable for unconstitutional conduct. Since 2009, "qualified immunity" statutes require that an officer's actions must violate a "clearly established law" (*Harlow v. Fitzgerald*, 457 US 800 [1982]). In practice, the requirement for "clearly established" law generally gives deference to officers (Michelman 2018).

Police Officers Are Incentivized to Get Home Safely, Clear Offenses, and Avoid Complaints

With the current lack of evidence on how hiring and training practices influence the pool of police officers, and in the absence of strong mechanisms to alter the incentives of police officers on the job, what do we know about what actually drives police officer decision making? We highlight three plausible candidates.

Staying Safe

Policing is an often-mundane job punctuated by bursts of extraordinary intensity and mortal peril, thus requiring a particular "hypervigilance" while on the job (Gilmartin 2002). Personal safety and survival are central and critical incentives guiding officer decision-making. Departmental policies will generally allow police to use force, even deadly force, when officers reasonably believe that their physical safety is threatened, or if officers reasonably believe that force is necessary to gain control of a situation.

Civilian encounters that are particularly high risk from an officer's perspective include traffic stops or domestic disturbances, where civilians may have access to weapons the officer can't see, or situations involving a civilian who is acting in a way that appears unpredictable to the officer (potentially due to a mental health crisis, substance use, or cultural differences) (Sierra-Arevalo 2021). Policy interventions that lower perceived threats to officer safety may reduce the frequency with which officers decide to use force. Such interventions generally receive strong backing from law enforcement organizations, if not the broader public, and include limiting the likelihood a civilian has a firearm or having trained health care professionals respond to situations involving someone experiencing a mental health crisis.[8]

Making Arrests

Agencies that use CompStat-style reviews in which individual officers are asked in front of their peers to justify the observed number crimes occurring in their beats create strong incentives for officers to provide a high level of engagement—in particular, to make arrests, issue citations, and lower crime (Sherman 1998). Conversely, failure to make arrests or issue citations is generally viewed as "depolicing" and adversarial to the department's interests (Prendergast 2001, 2021; Mas 2006; Shi 2009; Heaton 2010).

In 2019, approximately 46 percent of serious violent offenses and 17 percent of property crimes were cleared by an officer making an arrest, which is the required resolution for reporting a crime as "cleared" to the Federal Bureau of Investigation.[9] While it can be tempting to compare the "percent crimes cleared by arrest" over time or across agencies as a measure of officer productivity, this can be misleading for at least two reasons. First, a large fraction of arrests are made because the police observe a crime in progress, and these crimes in the "wide open" may be easier for police to deter than crime that occurs behind closed doors. Lower crime through increased deterrence would then reduce the average clearance rate (Cook 1979). Second, arrests may not be socially optimal in all situations. Not every public safety problem is best resolved with a criminal justice response, although this is the primary "hammer" in an officer's toolkit. Dangerous situations created by an individual in mental health crisis are an example. Police officers may be able to provide critical assistance in these situations. As government employees, officers frequently have better access to and information about other social services that are better suited to a specific crisis than a civilian would have (for example, police have direct communication with emergency medical services). But when the number of arrests, or arrests per crime, are the metrics used in supervisory meetings and published by the Federal Bureau of Investigation, officers are incentivized to appear active along these dimensions, rather than seeking to maximize the frequency with which they successfully connected someone with needed social services.

[8] For example, see the Firearms Policy Position Statement by the International Association of Chiefs of Police at https://www.theiacp.org/sites/default/files/2019-05/IACP%20Firearms%20Position%20Paper_2018%20(1).pdf, and in particular the recommendations for responses to mental illness.

[9] For details, see any version of FBI guidance to law enforcement agencies on completion of the Uniform Crime Reports or National Incident Based Reporting System: for example, the 2021 NIBRS manual at https://www.fbi.gov/file-repository/ucr/ucr-2019-1-nibrs-user-manua-093020.pdf/view.

Avoiding Complaints

Being the subject of a civilian complaint, even if it is ultimately not sustained, can create psychological stress for officers (Gilmartin 2002). To the extent that the chances of the complaint being sustained are small, as appears to be the case in Philadelphia and Chicago (Ba 2018), most of the actual utility loss for an officer occurs during the investigation process. Departmental level policies and practices, including the amount of unionization in an agency, can affect the intensity of this loss. How much of an officer's time is required during the investigation and are "real police work" activities otherwise limited? If a complaint is sustained, will the department require additional training, reassignment, suspension, or termination? Is there a possibility that the relevant Peace Officer Standards and Training commission will revoke a credential required for employment? Rivera and Ba (2019) document that the number of civilian complaints filed in Chicago significantly dropped after the police union notified officers of the seriousness of being involved in a complaint, which suggests that officers can and do respond to changes in the expected cost of complaints.

Exactly how the officers achieve those reduced complaints may or may not be welfare enhancing. Prendergast (2021) provides one way to think about this by modeling officer behavior towards crime suspects and crime victims, focusing on complaints as an outcome. A desire to avoid complaints can make officers less likely to use force against suspects but can also reduce the likelihood that officers will engage at all. What is critical is how likely a suspect subject to excessive force is to complain versus a victim who does not receive services. If the probability that a victim or suspect complains can vary with race and ethnicity, as suggested by Ba (2018) in racially segregated jurisdictions, an officer's incentive to avoid complaints will create spatial variation in how much police engagement occurs during any civilian encounter.

Conclusion

Police are the day-to-day physical embodiment of the government's coercive authority over its civilians. As such, police departments and officers are tasked with engaging civilians enough to reduce crime and promote feelings of safety, while not appearing to use force in an indiscriminate or biased way. This task is a complicated one, which officers are in some cases asked to solve in mere seconds, in uncertain and dangerous conditions.

The provision of any public good is likely to be suboptimal for some groups of individuals within a society. However, the context in which policing occurs means that the burden of over-provision of police engagement—that is, actions which make the general public feel safe, but may not reduce victimization by more than it reduces police legitimacy—will disproportionately be borne by minority groups. Variation in taste for police involvement is confirmed by a case study of complaints about over- and under- policing by racial and ethnicity of the complainant.

While police departments and officers are tasked with solving a complicated social welfare problem, the structure of institutional incentives is relatively simple. The dominant incentives faced by police departments are to develop policies which provide indirect benefits—to make civilians feel safe and to see the police "doing something" about crime. As long as only a small fraction of the population is directly affected by criminal victimization and only a small fraction of the population bears the cost of achieving the direct and indirect benefits of crime reduction, we would expect that rational, vote-maximizing politicians might not object to policies that either under-provided or over-provided police engagement in specific areas of concentrated disadvantage. The fact that crime rates, and the social cost of crime, rather than the frequency with which departments misallocate police engagement, are the primary metrics by which the police are publicly judged provides further incentives for crime reduction beyond what is socially optimal.

On top of truly optimal social policies being difficult to identify, individual officers within a department have substantial discretion in how they engage with the public. Standard strategies that organizations use to provide incentives for workers are limited by structured wage and promotion mechanisms, high monitoring costs, and limits on the ability to sanction employees. There is currently little evidence base with which one might identify screening, training, or monitoring strategies that support a department of officers who are able to make welfare enhancing decisions about civilian engagement.

With that in mind, we find that at least in large departments, more intensive screening of recruits, or more lengthy training requirements, is not obviously associated with the creation of a force that racially mirrors the civilian population, although there may be some scope for state-mandated training programs to actually reduce the diversity of the force. We also find little evidence that police departments attempt to compensate for relatively non-diverse police forces by providing more monitoring of officer behavior. Indeed, we also find that, in general, departments with more White officers relative to the patrolled population are less likely to formally solicit civilian feedback on training and monitoring procedures.

Given all of the structural challenges facing governments that want to provide public safety with police, it is not a surprise that civilian satisfaction with law enforcement is at a record low. The costs of police engagement as well as simply benefits, are becoming increasingly visible to departments and the public more broadly. The next challenge is to identify ways to incorporate those costs more explicitly into organizational and individual incentives of the police.

■ *Many thanks to Marion Aouad, Matthew Freedman, Anjelica Hendricks, Jacob Kaplan, Jenny Jiao, Philip McHarris, David Neumark, Canice Prendergast, Roman Rivera, Martha Stuckey, Naomi Sugie, and workshop participants at the Center for Population, Inequality, and Policy at University of California, Irvine, for comments on early versions of this paper. Oriana Ballardo provided excellent research assistance.*

References

Ang, Desmond. 2020. "The Effects of Police Violence on Inner-City Students." *Quarterly Journal of Economics* 136 (1): 115–68.

Ang, Desmond, and Jonathan Tebes. 2018. "Civic Responses to Police Violence." Unpublished.

Ba, Bocar A. 2018. "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago." Unpublished.

Ba, B., P. Bayer, N. Rim, R. Rivera, and M. Sidibe. 2021a. "Police Officer Assignment and Neighborhood Crime." Unpublished.

Ba, Bocar A., Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021b. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702.

Benoît, Jean-Pierre, and Juan Dubra. 2004. "Why Do Good Cops Defend Bad Cops?" *International Economic Review* 45 (3): 787–809.

Bentham, Jeremy. 1789. *The Principles of Morals and Legislation.* Franklin Classics.

Bindler, Anna, Nadine Ketel, and Randi Hjalmarsson. 2020. *Costs of Victimization.* Cham: Springer International Publishing.

Brayne, Sarah. 2014. "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* 79 (3): 367–91.

Brooks, L. 2008. "Volunteering to Be Taxed: Business Improvement Districts and the Extragovernmental Provision of Public Safety." *Journal of Public Economics* 92 (1–2): 388–406.

Bureau of Justice Statistics. 2018. "Census of State and Local Law Enforcement Training Academies, 2013." Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR36764.v1 (accessed September 15, 2021).

Chalfin, Aaron, and Justin McCrary. 2018. "Are US Cities Underpoliced? Theory and Evidence." *Review of Economics and Statistics* 100 (1): 167–86.

Cheng, Cheng, and Wei Long. 2018. "Improving Police Services: Evidence from the French Quarter Task Force." *Journal of Public Economics* 164: 1–18.

Cheng, Tony. 2019. "Input without Influence: The Silence and Scripts of Police and Community Relations." *Social Problems* 67 (1): 171–89.

Cohen, Mark A. 2014. *Costs of Crime.* https://onlinelibrary.wiley.com/doi/full/10.1002/9781118517390.wbetc204.

Comino, Stefano, Giovanni Mastrobuoni, and Antonio Nicolò. 2020. "Silence of the Innocents: Undocumented Immigrants' Underreporting of Crime and Their Victimization." *Journal of Policy Analysis and Management* 39 (4): 1214–45.

Cook, Philip J. 1979. "The Clearance Rate as a Measure of Criminal Justice System Effectiveness." *Journal of Public Economics* 11 (1): 135–42.

Cook, Philip J., and J. MacDonald. 2011. "Public Safety through Private Action: An Economic Assessment of BIDS." *Economic Journal* 121 (552): 445–62.

Cook, Philip J., Max Kapustin, Jens Ludwig, and Douglas L. Miller. 2017. *The Effects of COPS Office Funding on Sworn Force Levels, Crime, and Arrests: Evidence from a Regression Discontinuity Design.* Washington, DC: Office of Community Oriented Policing Services.

Di Tella, Rafael, and Ernesto Schargrodsky. 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack." *American Economic Review* 94 (1): 115–33.

Donohue, J. J., and S. D. Levitt. 2001. "The Impact of Race on Policing and Arrests." *Journal of Law and Economics* 44 (2): 367–94.

Draca, Mirko, Stephen Machin, and Robert Witt. 2011. "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks." *American Economic Review* 101 (5): 2157–81.

Durlauf, Steven N. 2006. "Assessing Racial Profiling." *Economic Journal* 116 (515): F402–26.

Evans, Williams N., and Emily G. Owens. 2007. "COPS and Crime." *Journal of Public Economics* 91 (1–2): 181–201.

Facchini, Giovanni, Brian G. Knight, and Cecilia Testa. 2020. "The Franchise, Policing, and Race: Evidence from Arrests Data and the Voting Rights Act. NBER Working paper 27463.

Fu, Chao, and Kenneth I. Wolpin. 2018. "Structural Estimation of a Becker–Ehrlich Equilibrium Model of Crime: Allocating Police across Cities to Reduce Crime." *Review of Economic Studies* 85 (4): 2097–2138.

Geller, Amanda, and Jeffrey Fagan. 2019. "Police Contact and the Legal Socialization of Urban Teens."

RSF: Russell Sage Foundation Journal of the Social Sciences 5 (1): 26–49.

Geller, Amanda, Jeffrey Fagan, Tom Tyler, and Bruce G. Link. 2014. "Aggressive Policing and the Mental Health of Young Urban Men. *American Journal of Public Health* 104: 2321–27.

Gilmartin, Kevin M. 2002. *Emotional Survival for Law Enforcement: A Guide for Officers and Their Families.* Casas Adobes, AZ: E-S Press.

Goncalves, Felipe, and Steven Mello. 2017. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–41.

Harris, Allison P., Elliott Ash, and Jeffrey Fagan. 2020. "Fiscal Pressures and Discriminatory Policing: Evidence from Traffic Stops in Missouri." *Journal of Race, Ethnicity, and Politics* 5 (3): 450–80.

Harvey, A., and T. Mattia. 2019. "Reducing Racial Disparities in Crime Victimization. Unpublished.

Heaton, Paul. 2010. "Understanding the Effects of Antiprofiling Policies." *Journal of Law and Economics* 53 (1): 29–64.

Heaton, Paul, Priscillia Hunt, John MacDonald, and Jessica Saunders. 2016. "The Short- and Long-Run Effects of Private Law Enforcement: Evidence from University Police." *Journal of Law and Economics* 59 (4): 889–912.

Hinton, Elizabeth. 2016. *From the War on Poverty to the War on Crime: The Making of Mass Incarceration in America*. Cambridge, MA: Harvard University Press.

Jácome, Elisa. 2020. "The Effect of Immigration Enforcement on Crime Reporting: Evidence from Dallas." https://elisajacome.github.io/Jacome/Draft_PEP.pdf.

Jones, Jeffrey M. 2020. "Black, White Adults' Confidence Diverges Most on Police." *Gallup*, August 12. https://news.gallup.com/poll/317114/black-white-adults-confidence-diverges-police.aspx.

Kaplan, Jacob. 2021. "Annual Survey of Public Employment Payroll (aspep) 1992–2016." Inter-university Consortium for Political and Social Research [distributor]. https://www.openicpsr.org/openicpsr/project/101399/version/V5/view (accessed September 25, 2021).

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203–29.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114 (3): 619–37.

Legewie, Joscha, and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." *American Sociological Review* 84 (2): 220–47.

Lerman, Amy E., and Vesla Weaver. 2014. "Staying out of Sight? Concentrated Policing and Local Political Action." *ANNALS of the American Academy of Political and Social Science* 651 (1): 202–19.

Levitt, S. D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review* 87 (3): 270–90.

Levitt, S. D. 2002. Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply. *American Economic Review* 92 (4): 1244–50.

Levitt, Steven D. 2004. "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not." *Journal of Economic Perspectives* 18 (1): 163–90.

Lum, Cynthia, Christopher S. Koper, Xiaoyun Wu, William Johnson, and Megan Stoltz. 2020. "Examining the Empirical Realities of Proactive Policing through Systematic Observations and Computer-Aided Dispatch Data." *Police Quarterly* 23 (3): 283–310.

Makowsky, Michael D., and Thomas Stratmann. 2009. "Political Economy at any Speed: What Determines Traffic Citations?" *American Economic Review* 99 (1): 509–27.

Makowsky, Michael D., and Thomas Stratmann. 2011. "More Tickets, Fewer Accidents: How Cash-Strapped Towns Make for Safer Roads." *Journal of Law and Economics* 54 (4): 863–88.

Makowsky, Michael D., Thomas Stratmann, and Alex Tabarrok. 2019. "To Serve and Collect: The Fiscal and Racial Determinants of Law Enforcement." *Journal of Legal Studies* 48 (1): 189–216.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *Quarterly Journal of Economics* 121 (3): 783–821.

McCrary, Justin. 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment." *American Economic Review* 92 (4): 1236–43.

McCrary, Justin. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *American Economic Review* 97 (1): 318–53.

Meares, Tracey L., and Emily Owens. 2019. "Third-Party Policing: A Critical View." In *Police Innovation: Contrasting Perspectives*, edited by David Weisburd and Anthony Braga. Cambridge, UK: Cambridge University Press.

Mello, Steven. 2018. "Speed Trap or Poverty Trap? Fines, Fees, and Financial Wellbeing. Unpublished.

Mello, Steven. 2019. "More COPS, Less Crime." *Journal of Public Economics* 172: 174–200.

Michelman, Scott. 2018. "The Branch Best Qualified to Abolish Immunity." *Notre Dame Law Review* 93 (5).

Miller, Amalia R., and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *Review of Economic Studies* 86 (5): 2220–47.

Muniz, Ana. 2012. "Disorderly Community Partners and Broken Windows Policing." *Ethnography* 13 (3): 330–51.

National Academies of Sciences, Engineering, and Medicine. 2018. *Proactive Policing: Effects on Crime and Communities*. Washington, DC: The National Academies Press.

Ornaghi, Arianna. 2019. "Civil Service Reforms: Evidence from US Police Departments." Unpublished.

Ouss, Aurélie, and John Rappaport. 2020. "Is Police Behavior Getting Worse? Data Selection and the Measurement of Policing Harms." *Journal of Legal Studies* 49 (1): 153–98.

Owens, Emily. 2017. "Testing the School-to-Prison Pipeline." *Journal of Policy Analysis and Management* 36 (1): 11–37.

Owens, Emily. 2019. "Economic Approach to 'De-policing'." *Criminology & Public Policy* 18 (1): 77–80.

Owens, Emily, David Weisburd, Karen L. Amendola, and Geoffrey P. Alpert. 2018. "Can You Build a Better Cop?" *Criminology & Public Policy* 17 (1): 41–87.

PAC. 2021. *Collaborative Review and Reform of the PPD Police Board of Inquiry: Policy Practice, and Custom Report*. Philadelphia: Police Advisory Commission.

Prendergast, Canice. 2001. "Selection and Oversight in the Public Sector, with the Los Angeles Police Department as an Example." NBER Working Paper 8664.

Prendergast, Canice. 2021. "'Drive and Wave': The Response to LAPD Police Reforms after Rampart." University of Chicago, Becker Friedman Institute for Economics Working Paper 2021–25.

Rivera, Roman G., and Bocar A. Ba. 2019. "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago." Faculty Scholarship at Penn Law 2109.

Rozema, Kyle, and Max Schanzenbach. 2018. "Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct." *American Economic Journal: Economic Policy* 11 (2): 225–68.

Rushin, Stephen. 2017. *Federal Intervention in American Police Departments*. Cambridge, UK: Cambridge University Press.

Schwartz, Joanna C. 2016. "How Governments Pay: Lawsuits, Budgets, and Police Reform." *UCLA Law Rev* 63.

Sharkey, Patrick. 2018. *Uneasy Peace: The Great Crime Decline, the Renewal of City Life, and the Next War on Violence*. New York: W. W. Norton Company.

Sherman, L. 1998. "American Policing." In *The Handbook of Crime and Punishment*, 429–541. Oxford: Oxford University Press.

Shi, Lan. 2009. "The Limit of Oversight in Policing: Evidence from the 2001 Cincinnati Riot." *Journal of Public Economics* 93 (1): 99–113.

Shjarback, John A., and Justin Nix. 2020. "Considering Violence against Police by Citizen Race/Ethnicity to Contextualize Representation in Officer-Involved Shootings." *Journal of Criminal Justice* 66.

Sierra-Arevalo, Michael. 2021. "American Policing and the Danger Imperative." *Law & Society Review* 55 (1): 70–103.

Strungis, D., K. Lyons, D. Kowalski, R. Cortez, N. Richman, L. Jones, and A. Parker. 2020. "Cities Shift Funds as Focus on Policing Grows: Credit Impact Hinges on Costs, Social Risks." Unpublished.

Webb, V. J., and C.M. Katz. 2014. "A Study of Police Gang Units in Six Cities." In *The Modern Gang Reader*, edited by Cheryl L. Mason, Arlen Egley, Jr., Jody Miller, and Malcolm W. Klein, 467–74. Oxford University Press.

Weisburd, David, and John E. Eck. 2004. "What Can Police Do to Reduce Crime, Disorder, and Fear? *Annals of the American Academy of Political and Social Science* 593 (1): 42–65.

Weisburd, Sarit. 2018. "Police Presence, Rapid Response Rates, and Crime Prevention." Unpublished.

Weisburst, Emily K. 2018. "Safety in Police Numbers: Evidence of Police Effectiveness from Federal COPS Grant Applications." *American Law and Economics Review* 21 (1): 81–109.

Weisburst, Emily K. 2019. "Patrolling Public Schools: The Impact of Funding for School Police on Student Discipline and Long-Term Education Outcomes." *Journal of Policy Analysis and Management* 38 (2): 338–65.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 29–48

Next-Generation Policing Research: Three Propositions

Monica C. Bell

Decades of social scientific labor have been poured into questions about policing and public safety, including a focus on uncovering the myriad racial disparities in policing and throughout the criminal system. Conversely, relatively little research has sought to disentangle policing from public safety. For example, previous research often failed to look at policing outcomes other than crime rates that are important for community well-being, or on how non-police alternatives could contribute to public safety. These research gaps have generated a bias in the research literature in favor of responding to concerns about public safety with status quo policing, perhaps with a few tweaks to existing training and practices.

But policing and public safety are not one and the same. Instead of starting from the presumption that more or better policing is the only route to public safety, many researchers are wondering whether organized community efforts could work better than traditional policing in achieving the goals of building public safety and improving community outcomes.

As one example of an alternative, Devone Boggan started the Peacemaker Fellowship in Richmond, California, as part of his work for the city's Office of Neighborhood Safety in 2009. In 2016, the Peacemaker Fellowship became Advance Peace, an organization that, among other things, runs an extensive mentorship and personal development program that engages individuals who are most at risk of engaging in interpersonal gun violence. The program offers these individuals 18 months of internships, travel opportunities, support navigating social

■ *Monica C. Bell is Associate Professor of Law and Sociology, Yale University, New Haven, Connecticut. Her email address is monica.bell@yale.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.29.

services, training and support for life goal-setting, and more and pays them for their participation. While the program engages in programmatic support for individual sustainability, some reports have characterized the work of Advance Peace and similar organizations as a form of friendship, highlighting its deep relational components. For example, Chabria (2017) reported that Boggan "has befriended 84 men considered most likely to kill with a gun" in Richmond (see also Bell 2019). Advance Peace also directly intervenes on episodes that could spiral into violence. The Advance Peace model has established programs elsewhere in California, such as Sacramento and Stockton and is gaining a foothold in Fresno (Hoggard 2020). However, Advance Peace does not work with law enforcement and does not hold building community trust in police as one of its goals. This promising program, which seeks to achieve safety and well-being by centering community, friendship, and support—not policing—has received scant attention from academic researchers.

As another example, Precious Blood Ministry of Reconciliation, a Chicago-based restorative justice organization, similarly builds relationships with young people, with additional focus on young people who have committed harm or who have had some relationship with the juvenile justice system. As an alternative to carceral approaches, some participants have received "Saturday Sanctions," in which the participant is engaged in self- and community-development programs (VanNatta and Kaba 2013). Today, the organization runs a battery of restorative programs, with focuses on several populations at particular risk of harm. The organization tells its own stories of impact by focusing on individual participants' experiences (https://www.pbmr.org/gallery#OurImpactStories).

In this essay, I offer three propositions for economists and other quantitative social scientists who aim toward a more accurate, comprehensive, and better contextualized research agenda for policing, public safety, and racial inequity: 1) the new public safety research must seek to probe the effects of policing on a range of outcomes, including education, health, community flourishing, and subjective well-being; 2) the new public safety research should evaluate policing alternatives, including community-based strategies of crime deterrence and accountability for harm; and 3) the new public safety research should reach beyond the study of racial disparities to investigate the effects of racism on crime, harm, and disparity in the criminal legal system. In short, the next-generation research agenda on policing and public safety must respond to the deficiencies and limited focus of past research and chart a less parochial course.

Proposition 1: The new public safety research must seek to probe the effects of policing on a range of outcomes, including education, health, community flourishing, and subjective well-being.

The traditional datasets used by economists fail to measure many of the collateral costs of policing. The bulk of research on policing and public safety has focused on crime rates (especially felonies), fear of crime, and internal police department dynamics (stops, arrests, clearance rates, response times, expenditures, and police force size) (Chalfin and McCrary 2017; Raskolnikov 2020). Because of the multiple functions of police in society, and because achieving safety requires more than an exclusive focus on lowering crime rates and making police departments more efficient, scholarship focused on policing and public safety should measure a broader set of outcomes than are usual in this field. The next generation of police research should examine the full range of relationships between policing and other aspects of community life and social structure—including effects on health, segregation, education, urban development, municipal budgets, wealth consolidation, labor relations, liability insurance, and other metrics. Police deterrence of crime, even if successful, may have collateral costs for other metrics of concern.

Some quantitative sociologists have tackled these outcomes, with important implications for conceptual understandings of the consequences of policing beyond crime levels. For example, Legewie and Fagan (2019) examined how aggressive policing, specifically "Operation Impact" in New York City, affected the educational performance of Black and Latinx youth between 2003 and 2012. Legewie and Fagan find a causal relationship between harsh policing and lower educational performance. They use variation in the timing of police surges across neighborhoods and a difference-in-differences approach to show that exposure to surges of police hyper-presence in New York City neighborhoods leads to lower test scores among Black male adolescents, with increasing effect sizes by age. Building from these and other empirical insights, Justice (2021) recently theorized that the effects of policing and incarceration on the educational opportunities and experiences of American children create a form of "hobbling," or "a social process by which the massification of policing and incarceration systematically compromises the ability of target demographics of American children to enjoy their rights to a free and appropriate public education" (3.1).

American policing and other criminal system researchers have traditionally relied upon five national databases of large-scale secondary data in their analyses of criminal system outcomes: Uniform Crime Reports (UCR), National Incident-Based Reporting System (NIBRS), National Crime Victimization Survey (NCVS), Law Enforcement Management and Administrative Statistics (LEMAS), and Law Enforcement Officers Killed and Assaulted (LEOKA). Researchers have offered a number of critiques of these datasets, mostly focusing on aspects of underreporting and undercounting. For example, UCR offers a wealth of information about reported crime in the United States; however, it is plagued by a serious missing data problem because many police departments do not comply with requests to submit their crime reports to the Federal Bureau of Investigation. This risk seems greatest among police departments in locations where crime might be higher than average (Lynch and Jarvis 2008; Boylan 2019). Similarly, researchers have criticized data from the NIBRS for unrepresentativeness and potential underestimation of crime rates (McCormack, Pattavina, and Tracy 2017). The NCVS, which helpfully allows a deep dive on experiences of victimization to some degree—including otherwise unreported crime—also comes with important limitations such as declining response rates, the omission of crimes against youth under age 11, underestimation of serial victimization because of rules that sometimes collapse multiple victimizations into a single incident, likely underestimation of rape and sexual assault, and other issues.

These narrow ranges of research and data flow from researchers' presumption that the central function of policing is to reduce and respond to crime. However, decades of research have also shown that policing has played many other latent roles in the American social order. For example, policing has functioned to confine and control socially disfavored groups, both in early American history and in the present (for example, Harcourt 2007; Wacquant 2009), and to protect the property of wealthy groups, including tamping down labor movements seeking fair wages and hours (Fisk and Richardson 2017; Levin 2020). In my own scholarship, I have written about how police policies and practices function to shape families' residential preferences (Bell 2020a) and in this way, reinforce racial residential segregation across cities and suburbs (Bell 2020b; see also Fagan and Ash 2017; Gordon 2020; Kurwa 2020).

Other scholars have extended a similar lens to the mental and physical health outcomes of policing, both for those who directly bear a criminalized status and indirectly for members of their families and communities (Asad and Clair 2018). For example, Sewell and colleagues have found numerous associations between harsh, surveillance-style policing and negative health outcomes. They find associations between chronic stop-and-frisk at neighborhood levels and emergency room use (Kerrison and Sewell 2020); between exposure to lethal police violence and chronic illness like high blood pressure and diabetes (Sewell et al. 2021); and between living in predominantly Black neighborhoods with heavy policing or predominantly White neighborhoods with racially targeted heavy policing, and various negative health outcomes (Sewell 2017). Other examples have emerged, especially in the most recent decade: for example, Geller et al. (2014) provide evidence on associations in survey data of young men in New York City between symptoms of trauma and anxiety and the number of police stops these men had experienced, together with their perceptions of the intrusiveness and fairness of these stops.

Researchers can also turn to new sources of data to investigate community flourishing, community power, and individual well-being. A growing body of scholarship, primarily in positive psychology, aims to develop metrics for individual well-being and community flourishing. For example, the Human Flourishing Program at Harvard University's Institute for Quantitative Social Science has developed a twelve-question survey, along with some additional context-specific tools, for measuring individual flourishing (Harvard IQSS 2021). The twelve questions are intended to reach six central domains of human flourishing: happiness or satisfaction, health (mental and physical), an individual's sense of purpose, an individual's character and virtue, the presence and nature of close interpersonal relationships, and financial/material stability (VanderWeele 2017, p. 8149). VanderWeele (2019), the primary theorist of this particular human-flourishing measure, has also developed potential measures of community flourishing, building from six domains: individual flourishing (to be measured using the aforementioned twelve items), positive relationships, good leadership, healthy community practices, community satisfaction, and a sense of community mission. He argues that the other five domains could be measured through an additional 20-item questionnaire (VanderWeele 2019, pp. 258–60). Researchers aiming to recognize the latent functions of policing through their research practices might incorporate some of these individual- and community-flourishing measures into the evaluation of policing, police reforms, and public safety-oriented state and community interventions.

Other scholars are also developing frameworks and tools that form the basis of an alternative set of metrics for policing and public safety outcomes. For example, Sampson (2012) reports evidence from the Project on Human Development in Chicago Neighborhoods, which gathered data through surveys, interviews, data on neighborhood physical conditions using video technology, and responses to events, and combined it with other available data on health, crime, housing, violence, and population. Sampson draws from that data to develop measurable concepts such as collective efficacy, legal/moral cynicism, and other key outcomes that should be of central concern in studies of policing. Cohen et al. (1998) discuss how to use surveys like the Stress-Related Growth Scale and the Posttraumatic Growth Inventory to measure "thriving," by which they mean the response to a specific stressful event. They also discuss how to validate self-reported data and how to carry out such assessments for groups and communities. Small (2009, 2017) offers examples of how to estimate the size and strength of social support networks and other measures of social capital, important potential outcomes for studies that might explore how policing affects the expansiveness and strength of social bonds. A variety of authors have looked at relationships between policing and measures of voting or civic engagement. For example, Drakulich et al. (2017) find that being stopped by police tends to increase voting, but experience of incarceration tends to decrease it. Walker (2020) combined data from different national surveys to argue that the experience of having a loved one or family member who has contact with the criminal system can increase voting, mobilizing people to use their political voice to change political and legal conditions. Laniyonu (2019) finds that concentrated policing across areas of New York City "was associated with reductions in voter turnout in the 2006 and 2010 midterm elections, it was associated with higher rates of turnout in the 2008 presidential election . . ." This insight offers suggestive evidence of a combination of chilling effects and mobilizing effects on voters that varied depending on the role of policing in local political debates.

It is important for researchers to embark on this expanded body of quantitative work with a rich understanding of the theories that would link policing to these outcomes in other domains. As one example, it may not make sense to expect significant changes or variation in civic engagement or collective efficacy related to a single intervention in a year or less, given the depth of legal estrangement and marginality in communities studied.

Proposition 2: The new public safety research should evaluate policing alternatives, such as community-based strategies of crime deterrence and accountability for harm.

Economists, like other social scientists, rarely research community-based crime deterrence efforts. This empirical neglect biases the evidence base that informs policy debates. Past research on public safety across social science disciplines has overwhelmingly focused on the capacity (or lack thereof) of police to deter crime and reduce crime rates. We are now a generation into a body of research that has come to treat the capacity and necessity of police to prevent crime as a truism (Meares 2014; Sharkey 2018). Indeed, the question of whether police can and do prevent crime is so well-worn that it is no longer treated as a research question but as a fact. That fact has given birth to an array of investments into policing that focuses on deterrence rather than response, including a multitude of for-profit companies that provide surveillance and predictive algorithms that assist police in crime prevention (Brayne 2020; Ferguson 2017). It is worth pausing a moment to reflect on how we have arrived at this point.

For much of the 20th century, quantitative research failed to find a connection between police and crime prevention (Meares 2015). In the early to mid-1990s, scholars invested serious energy debunking "the myth of the police," arguing that "[t]he police do not prevent crime . . . Experts know it, the police know it, but the public does not know it" (Bayley 1994, p. 3). Much of the 20th-century policing research didn't examine policing strategies aimed at reducing crime before it occurred, but policing strategies in response to crime that had already occurred. For example, in a nod to the baseline 20th century project of professionalizing police forces and solidifying an occupational identity and purpose among police officers, much policing research examined the structure and efficiency of departments (for example, Reiss 1992). Research on these topics continues in the 21st century (for example, Cihan, Zhang, and Hoover 2012; Vidal and Kirchmaier 2018). Yet now, an expansive body of research on algorithmic policing and crime deterrence suggests, for example, that targeted policing in more condensed "hot spots" is an effective preventive policing strategy (for example, Braga, Papachristos, and Hureau 2014). More generally, preventive policing, once deemed impossible, has become a universally accepted aspiration of police departments.

Since the late 1990s and early 2000s, in the wake of the "Great Crime Decline" of the 1990s, scholars have been trying to sort out how substantial and central a role policing played in that drop in crime. In this journal, Levitt (2004) gives his evaluation of ten possible factors, claiming that increased numbers of police played a determinative role in the 1990s decline in crime, but better policing strategies did not. Some book-length treatments of reasons behind the Great Crime Decline include Blumstein and Wallman (2000), Roeder, Eisen, and Bowling (2015), and Sharkey (2018). They all offer considerable nuance but also acknowledge ongoing uncertainty about the precise reasons for the crime decline. A number of studies provide suggestive evidence that various approaches to policing were among a number of factors in the 1990s crime decline, along with others like the rise and fall of the crack epidemic and the density of community organizations. By the early 2000s, some economists advocated uncritically for putting more police on streets, based on their belief in a straightforward causal relationship between the flooding of cities with more police on the streets in the 1990s and reduced levels of crime (for example, Ludwig and Donohue and 2007). They neglected the racialized costs of this strategy, as they did not account for outcomes like broken community bonds, loss of interpersonal and institutional trust, or other social and political costs of heavy, concentrated policing.

To be sure, it is plausible that during the 1990s crime decline, increases to police funding and new methodologies of policing contributed to reduced crime—while also causing myriad collateral harms. Moreover, as sociologist Patrick Sharkey (2018) has shown, local community organization presence also had a causal relationship with those declines during the 1990s and 2000s in cities throughout the United States (Sharkey 2018; see also Sharkey, Torrats-Espinosa, and Takyar 2017). Sharkey's work uses the formation of nonprofits focused in other areas, like arts and humanities, medical research, and environmental protection as an instrumental variable for the formation of nonprofits related to violence, crime, and community-building, thus allowing a causal estimate. Surprisingly little social scientific research in sociology, criminology, or economics has similarly focused on the role of community-based organizations in crime reduction or community-based alternatives to policing and prisons. Part of the reason is that there has never been a robust, well-funded, and consistently supported network of community organizations that engage in violence reduction and public safety efforts, which might leave the impression that policing must be the primary way to reduce violence, while blinding researchers and analysts to the capacity for community organizations to play a more central role if they were better and more consistently funded, supported, and evaluated.

The primary alternative to policing that has received thorough attention from researchers are violence interruption programs, such as CureViolence, CeaseFire, Safe Streets, Save Our Streets, and hundreds of others, which enlist the work of trusted community members to interrupt encounters that could become violent without their support and negotiation. The basic idea of this violence prevention approach is to treat violence before it occurs by detecting and mediating conflict before it escalates and to do this with community members who have a type of expertise, skill, and credibility that police officers lack. These violence interruption programs have shown an array of promising results for reducing violence over more than a decade of research (Abt 2019; Braga and Pierce 2005; Milam et al. 2013; Slutkin et al. 2015; Webster et al. 2013; Whitehill et al. 2014).

But there are a number of other community-based programs or alternatives to traditional policing that remain largely unstudied, even though some of them are becoming models for other jurisdictions across the nation. For example, CAHOOTS (Crisis Assistance Helping Out on the Street) started in Eugene, Oregon, to send two-person clinical response teams to aid people in mental health crisis, without relying on armed police officers. Although the program has existed for more than three decades, in summer 2020 it gained national attention and became the model for numerous pilot programs—in San Francisco, Denver, Rochester, Toronto, and more. Eugene's CAHOOTS program is funded and overseen by the police department, but some other emerging programs are funded and managed separately from police. Despite its long duration—even longer than the violence interruption programs mentioned above—CAHOOTS has never been rigorously evaluated. There are also rich debates over, among other things, how to measure its diversion rate (Gerety 2020). There is a dearth of information and modeling of police-free crisis response, though one hopes that will change as more cities embrace these approaches.

Similarly, little rigorous research examines the effects of interventions targeted directly at conditions that produce crime. One example is Advance Peace, described in the introduction, which aims to tackle the lack of career opportunities and economic resources that drive participation in crime. Chicago's Rapid Employment and Development Initiative (READI) program offers one promising approach for intervening in the criminogenic conditions of poverty, housing insecurity, low economic opportunity, and trauma while laying groundwork for empirical measurement of this approach. READI works with six community organizations to offer participants who are people deemed at risk of participation in violence one year of transitional employment, cognitive behavioral therapy, and professional development. READI partnered with the University of Chicago Crime Lab early and is operating as a randomized control trial to measure an array of program outcomes. While there is some research on the effectiveness of restorative and transformative justice projects that aim to respond to and heal individuals and communities after episodes of violence without using prisons and police, more is needed to help these programs be effective as they evolve and scale up (Dixon 2020; Sered 2019).

Some local governments have expanded their public safety and crime reduction efforts beyond policing as well. For example, the New York City Mayor's Office of Criminal Justice has pioneered "NeighborhoodStat," a community-led strategic plan to address specific social, economic, and environmental conditions affecting crime across 15 public housing developments (Pearl 2019). NeighborhoodStat has shown some potential effects on crime, though the plan and the evaluation process are still in relatively early stages (Delgado et al. 2020).[1] "NeighborhoodStat" is a reference to a famous data-gathering effort, CompStat, pioneered in New York City in the mid-1990s under Commissioner William Bratton, which is now used in many police departments nationwide (Weisburd et al. 2003). CompStat measures complaints about crime from the public, police arrests, and issuance of summonses. Most police departments using this approach hold weekly meetings, open to the public, where officers learn about and reflect upon the CompStat numbers as a police management strategy. Though CompStat was initially lauded (for example,

[1] The author has served as a paid member of an advisory committee for the evaluation of this program, which is part of the Mayor's Action Plan for Neighborhood Safety.

O'Connell 2001), it has faced recent criticism both because of the limited scope of data collected and because it incentivizes officers to search, engage, arrest, and perhaps to distort crime statistics (Bronstein 2015; Lu, Yang, and Thomas 2020; Thomas and Wolff 2020). In addition, CompStat is run by the police department and focuses on evaluation of police. In contrast, NeighborhoodStat is run by the Mayor's Office in collaboration with the community and aims to evaluate a complex set of alternative investments in public safety, ranging from increased outdoor lighting to exercise programs to youth employment opportunities.

Okechukwu's (2021) research on community-based safety measures in mid-20th century Brooklyn provides one example of the potential for an expanded scope of research on community-based safety measures. Okechukwu probes archival materials and oral histories to describe four different non-police organizational strategies that Black community members in the Brooklyn neighborhoods of Bedford-Stuyvesant and Crown Heights used to produce public safety: externally focused community patrol, internally focused community patrol, building physical refuge for hyper-vulnerable populations, and "othermothering" (older women serving as caring "eyes on the street" to observe and keep communities abreast of threats) (see also Collins 2002; Jacobs 1961 [2016]). Okechukwu analyzes each strategy, exploring how well (if at all) they fit with contemporary visions of abolition of the police. This work exposes both the possibilities and limitations of community-based security measures. While Okechukwu's work is qualitative, quantitative research could also have value for exploring other aspects of these questions, such as the relative long-term efficacy of different longstanding community-based strategies as compared with each other.

One Million Experiments, a collaborative project between two police and prison abolitionist organizations, Project Nia and Interrupting Criminalization, is collecting snapshots of community-led safety efforts across four categories—mutual aid, alternatives to calling 911 during emergencies, support (such as healing circles and financial support), and community events (One Million Experiments 2021). The project also produces a newsletter that does periodic deep dives on particular community safety projects. One Million Experiments does not seek to evaluate these projects, but to expand the imaginations and agendas of readers about projects they could attempt in their own communities.

These examples are seeds for economic research on non-carceral approaches to violence reduction and response to harm. Scholars should examine the mechanisms and characteristics of organizations that can affect crime reduction. Perhaps some types of crime are better deterred by alternative organizations than by traditional police. Perhaps increases in crime in the short term would be offset by crime reductions in the longer term. Perhaps there would be heterogeneous effects of organizations on crime and well-being depending on neighborhood and municipal characteristics. Scholars should also continue to investigate how non-carceral ecological interventions might reduce crime, building on preexisting research on greening, street lamps, and more (Doleac and Sanders 2015; Garvin, Cannuscio, and Branas 2013).

At a basic level, what sociologist Robert Sampson (2012) has termed "collective efficacy" may account for how the presence of community organization seems to produce lower violent crime rates. A community's sense of mutual trust and informal social control over what happens in their neighborhoods can perhaps build protective factors that soften the criminogenic effects of social and economic neighborhood disadvantage. Indeed, rather than passively accepting the underlying assumption that police departments and policing techniques in more-or-less their current form are a natural and inevitable feature of a public safety agenda, researchers should investigate of the accuracy of those assumptions. If the policy goal is improved public safety along with other aspects of community flourishing more generally, community-based organizations may well have a much larger role to play.

Proposition 3: The new public safety research should reach beyond the study of racial disparities to investigate the effects of racism on crime, harm, and disparity in the criminal legal system.

Economists arrived relatively late to the study of racial disparities in the criminal system compared other social sciences (for example, compare Becker 1968 with Du Bois 1904; see also Bushway and Reuter 2008). Yet once economists entered the fray, they began to have great influence over the methods, assumptions, and outcomes of interest in policing and public safety research. Economists have explored racial disparities in policing along multiple dimensions in papers that have guided the way for other quantitative social scientists who attempt to isolate discrimination as a cause of racial disparity at particular junctures in the criminal system continuum (for example, Antonovics and Knight 2009; Anwar and Fang 2006; Coviello and Persico 2015; Goncalves and Mello 2021). Economists have also shed light on the effects of police officer race on racially disparate outcomes (for example, Donohue and Levitt 2001). Some research has focused instead on structural features, such as urbanity and racial demographics, to understand racial disparities in policing (Hoekstra and Sloan 2020; Ross 2015).

In some permutations, this style of research has laid groundwork for deeply consequential changes in law and policy meant to address those racial disparities: for example, it can be useful to know the details of how racial disparities in traffic stops operate to identify particular ways to reduce those disparities. However, this style of research on racial disparities—focused as it is on specific types of interactions with the criminal justice system, as reported by the criminal justice system—is routinely oblivious to the social, political, and economic context of race. Economists' analytic approaches to studying racial disparities lead them to miss many of the institutional mechanisms through which racism arises in the criminal legal system (as discussed in this journal by Small and Pager 2020). To convey that context, sociologists sometimes use the term "structural racism," which can be defined as "a social system in which race is a central principle of social organization that serves to sort individuals into positions of relative advantage and disadvantage based on their racial category" (Merolla and Jackson 2019: 2; see also Gee and Ford 2011; Powell 2008).

The interpretation of quantitative results of racial disparities focusing on specific steps in the policing process has led to some contentious debates over claims that a given set of police data does or does not reveal officer racial bias. Researchers should be much more attentive to the limitations of their data, their choice of a theoretical lens through which they interpret their data, and the pressures to report results on racial disparities in a provocative, iconoclastic way.

As one example, quantitative psychologists Johnson and colleagues published an influential 2019 article in the *Proceedings of the National Academy of Sciences*, which stated (in the article's statement of significance): "White officers are not more likely to shoot minority civilians than non-White officers" (Johnson et al. 2019, p. 15877). After criticism for making this claim based on the data and estimation strategy used (Knox and Mummulo 2020), and after a back-and-forth in letters, Johnson and colleagues issued a correction and reframed their finding in this way: "As the proportion of White officers in a fatal officer-involved shooting increased, a person fatally shot was not more likely to be of a racial minority" (Johnson et al. 2020a p. 9127). Although this correction stated their core finding in a more careful if perhaps less publicly digestible way, Johnson and colleagues ultimately retracted their article altogether because they believed that their "work has continued to be cited as providing support for the idea that there are no racial biases in fatal shootings, or policing in general" (Johnson et al. 2020b, 18130).

Perhaps the most well-known social science controversy on interpreting data on racial disparities in policing involves the research by Fryer (2019). Fryer's research concludes, based on data from ten major cities, there are racial disparities in officers' uses of physical force like handcuffing, pepper-spraying, and other non-lethal engagement, but there was no racial disparity in the likelihood of being shot by police. As commenters pointed out when working-paper versions of the study became available, these data were on people who had a police encounter. Thus, key aspects of racial inequality in uses of force, such as the higher rates at which Black people are stopped by the police in the first place (Knox, Lowe, and Mummolo 2020), as well as other tricky aspects of studying police interactions through data are not taken into account (Goff et al. 2016). Accordingly, the study cannot shed light on a central pathway through which racism affects policing (for example, Epp, Maynard-Moody, and Haider-Markel 2015). Fryer (2020) followed up in a June 2020 op-ed in the *Wall Street Journal.* After expressing "dismay" because many readers have misinterpreted or misappropriated this finding "as evidence that there is no racism in policing, that football players have no right to kneel during the national anthem, and that the police should shoot black people more often," Fryer explained that from his viewpoint as an economist, his study cannot speak to the impact of racism on its outcomes: "Racism may explain the findings, but the statistical evidence doesn't prove it. As economists, we don't get to label unexplained racial disparities 'racism.'"

Of course, there are better ways to address the problem of unexplained racial disparities than leaving them untheorized. Researchers can study policing outcomes through a lens that employs theoretically informed metrics for racism. Part of what it means to study racism is to draw upon substantive theories on race and the law. For example, Rojek, Rosenfeld, and Decker (2012), examined racial disparities in the traffic stops in St. Louis, Missouri. They observed interaction effects of officer race, driver race, and the racial composition of the neighborhood: In predominantly White neighborhoods, stops of Black drivers were more likely to result in a search, especially when the officer was White. However, in predominantly Black neighborhoods, White officers were more likely to search cars with White drivers. The researchers did not merely report this finding; they drew upon Donald Black's theory of law and suggested that both outcomes on racially disparities are the consequences of both racial profiling and residential segregation.

Health researchers have been at the forefront of moving beyond just providing evidence on racial disparities in outcomes, and are starting to examine associations between health and racial disparities that themselves are a legacy of racially unjust institutions (Asad and Clair 2018; see also Kohler-Hausmann 2019). As one of an increasing number of examples, Boen, Keister, and Aronson (2020) examine associations between aspects of the racial wealth gap and health outcomes. They find that "savings, stock ownership, and homeownership consistently improve health, but debt is associated with worse health, even after adjusting for total net worth." They also find that the correlations between different kinds of financial assets and health vary by race. In another study, Boen, Kozlowski, and Tyson (2020) look at correlations in school-, family-, and individual-level data between indicators of perceptions of safety in schools and health outcome both in adolescence and after leaving school in young adulthood. Lukachko, Hatzenbuehler, and Keyes (2014) consider state-level indicators of racial disparities in employment, education, political participation, and sentencing treatment by judges (including incarceration, capital sentencing, and disenfranchisement) to examine the effects of structural racism on heart attacks. In this style of work, we see possibilities of studying the effects of racism on health by examining the effects of a central outcome of structural racism, such as wealth inequality or educational disadvantage, and we can also imagine creating a multifaceted measure that aims to capture an array of aspects of racial marginality and examining its effects on health.

Researchers can also study public safety outcomes through a lens that employs theoretically informed metrics for structural racism. Drawing from this work, the next generation of quantitative research on policing and public safety needs to account for structural racism, not only in interpretation of research outcomes, but also as it manifests in the data itself.

This type of critical eye is also needed for exploring racism within police organizations, especially police culture and networks. The groundbreaking work of Wood, Roithmayr, and Papachristos (2019) use social network analysis of "big data" not to study police-generated crime data, but instead to study the officers themselves, their networks, and police misconduct complaints. They find that police misconduct in

Chicago is not random, but networked, meaning that officers tend to engage in behavior that leads to a misconduct complaint in groups. They also find associations between misconduct incidents and police demographics, such as age and race—younger officers receive more civilian and departmental misconduct complaints than older officers, and White officers are somewhat more likely than Hispanic and Black officers to have received at least one complaint (p. 13). Studying networks may help quantitative sociologists to get a handle on dynamics that influence culture within organizations, including the networked quality of police misconduct. Police culture is sustained by an array of professional narratives: narratives of dangerousness (Sierra-Arévalo 2019), narratives of "bad" segregated neighborhoods (Moskos 2008), and so forth. Quantitative researchers should explore opportunities to operationalize these narratives in understanding how racism produces racial disparities in policing.

Next Steps

While this article maintains that increased research and evaluation of alternative projects would be germane to policy debates, it is important to use the tools of social science with circumspection and humility. It may be that some norms of "evidence-based policymaking," which base normative decisions about good policy on clear, countable results, are in some ways out of step with creative efforts to "reimagine" public safety. Evidence-based policymaking, at least as currently conceived, is often backward-looking and timid. Reimagination is forward-looking and definitionally bold. This difference may come into play in the evaluation of community-based projects. Inevitably, many of these programs, at least initially, will be riddled with mistakes—perhaps even fail (for example, Madden, Leeds, and Carmichael 2020). This process of trial and error is at the heart of reimagination. As Ejeris Dixon (2020, 19) has explained, "We have to be accountable enough to continue our experiments, to measure them, to hold ourselves to high standards, and to believe in them."

After all, embarking upon experiments for societal improvement is neither utopian nor unprecedented. American history is littered with institutions that were once deemed abject failures that might have been discarded if not for belief in them beyond data. Ironically, police departments themselves can be viewed as an emblematic example of such evolution. Early police departments were often initially riddled with corruption and unprofessional. Until research in the 1990s, researchers were unable to provide a persuasive statistical argument that the police reduced crime. It was only after a radical rethinking of police work—not constrained by preexisting evidence—that new realities could emerge.

Thus, researchers who embark upon collaborative research on community-based safety projects must be careful about how to interpret what may appear to be "failure." Statistical failure may not mean that the project is fundamentally valueless. Along similar lines, some community organizers have criticized traditional quantitative research paradigms for deploying research strategies that misunderstand the complexity of organizations' goals and desired impact, instead externally imposing logics of traditional measurement where alternative evaluative logics, perhaps based on narrative, might be more valuable (for example, Keene, Keating, and Ahonen 2016; Roe 1994; Rogers 2008; van Wessel 2018). Qualitative evaluations that draw from participant narratives have allowed researchers to understand outcomes of programs and policies that were not previously anticipated or that are difficult to measure, to explore unanticipated mechanisms of how and why approaches succeed or fail, and to explore nuances that quantitative tools cannot capture. Qualitative evaluations can also give voice and power over data-gathering to communities themselves, serving broader democratic goals and rectifying epistemic injustices (or wrongs against people "in their capacity as a knower" (see Fricker 2007, p. 1), which can be potent for marginalized communities in criminal system policymaking processes.

Beyond using mixed-methods research, both qualitative and quantitative, in evaluation—a strategy that has become standard—some researchers have gone further by embracing community-based participatory research as part of rigorous evaluation processes, enlisting the collaborative effort of both qualitative researchers and community members. The belief is that the data are better analyzed, with fewer framing and interpretation risks of the sort outlined above, by triangulating multiple types of data and by staying tightly connected to the population being studied throughout the research process. Social scientists should approach evaluative research with awareness that, while quantitative research is a valuable tool, it should never be an exclusive tool in moral and political debates over public safety.

Over the past few decades, academic public safety researchers have been much more aggressive about partnering with police departments than with community organizations. Social scientists tend to receive esteem for reaching data-sharing agreements with governmental agencies, including police departments, that allow the researchers to clean and analyze administrative data. Leading social science journals often publish and prominently feature articles that emerge from academic–police department partnerships. The esteem directed toward scholarship that emerges from such partnerships comes in part because of the increasing influence of "big data," both within policing and in the study of it (for example, Brayne 2020; Desmond, Papachristos, and Kirk 2016). Perhaps the quintessential example of such partnership is the storied Chicago Crime Lab, mentioned above, which has roots within the Chicago School of Sociology that was famous for treating the city as a laboratory.

Of course, analysis of police data can be of high value for understanding racial disparities, police budgets and staffing, and other aspects of policing. However, research that uses police data has severe limitations. Especially with respect to crime data, police data tell scholars about the world as it is understood and created by the institution that collects the data. Much crime goes unreported. Police presence is not equal across space, and in some places—especially White and suburban communities—crime does not consistently receive a police response (Jacques and Wright 2015). A current University of Chicago sociologist Robert Vargas (2020) has offered a trenchant critique of the Chicago Crime Lab for its history of research partnerships and deep entanglements with the Chicago Police Department. In Vargas's view, such partnerships have led to flawed research that seeks to "improve cities by managing Black individuals instead of ending the police violence Black communities endure." For these and other reasons, research that relies on police-created data is not more valuable, more scholarly, or more deserving of funding and attention than research on community-based modes of deterrence and accountability.

Pursuing research on alternatives to traditional, carceral crime deterrence measures will also demand that researchers engage in new forms of partnership. There are numerous community-based efforts afoot to increase community security and address harm, but most public safety researchers treat these efforts as if they are invisible or unworthy of rigorous study. It should be admitted that, in the past, the professional rewards for academics who consider pursuing empirical research on community-based alternatives to policing and public safety have often been scant. Quantitative scholars might be hesitant to make the time and energy investments to pursue a partnership with community-based organizations in part because their data, if the organizations have it, require working with smaller sample sizes. Research produced in partnership or collaboration with community organizations may be labelled as "activist" scholarship, perhaps even imposing career costs for its practitioners (Cancian 1993; Warren et al. 2018). However, collaborating with community groups is important work that the academy has the capacity to increase. Examples like the aforementioned READI program, with its evaluative collaboration with University of Chicago researchers, support this point. Yet, much more of this work is needed outside of the large university research labs. Scholars in the academy, including those who review articles, evaluate grant proposals, and vote on faculty appointments, should reflect upon and reorient our assumptions about what constitutes a laudable institutional partnership and rigorous scholarly research. Funders, including the federal government, might also have a role to play in facilitating this research by funding the alternative harm prevention and response work of community groups in ways that would better facilitate data-gathering and causal identification.

The current political and moral conversation over public safety is shallow and truncated because it is impossible for policymakers and activists to have well-informed debates about costs and benefits, tradeoffs, predicted outcomes over time, or potential unintended consequences of various changes. For example, if pursuing alternatives to policing along with a corresponding decrease in police funding would produce a sharp increase in homicide rates, the utility calculus might be different than if these approaches would produce a slight uptick in property crime in wealthy areas while increasing well-being in marginalized areas.

In addition, some recent reductions in police spending happened for reasons unrelated to changes in public safety policy. The city of Oakland, California, has been used as an example of the perils of cutting police funding and investing in alternatives—even though the 2020 police budget reduction in Oakland was taken up as an austerity measure because of municipal fiscal crisis, not to embrace a political goal of reimagining public safety. Indeed, the Oakland police budget reduction occurred alongside a reduction in the fire department's budget, and some line items cut from the police budget were actually cuts from alternative measures to promote public safety that were funded through the police department, such as Operation Ceasefire mentioned earlier (Clayton 2021; Sciacca 2021). Research is sorely needed to compare police defunding when it is accompanied by investment in various alternatives and when it is not so accompanied.

This essay has offered three propositions for next-generation research on public safety, which is already now underway. In a process accelerated in the wake of the murder of George Floyd in June 2020, social scientists have finally started to question the role of our own research in perpetuating the status quo, our uncritical acceptance of professional tropes about the value of police data, our ideas about the best ways of measuring public safety, and more. This paper offers an entrée into a broader plan for a social science that not only sees those status quo biases in our work, but actively seeks to root them out. Perhaps most importantly, next-generation policing and public safety research must remain circumspect and humble. Our expertise, while vast in some ways, is limited in others (Simonson 2021). Sometimes, the data will simply be insufficient to provide complete answers to the burning questions of our day. At those times, we must pause and reflect on where our findings fit within a larger ecosystem that is examining and rethinking policing, the criminal system, and their transformation.

References

Abt, Thomas. 2019. *Bleeding Out: The Devastating Consequences of Urban Violence--And a Bold New Plan for Peace in the Streets.* New York: Basic Books.

Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91 (1): 163–77.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–51.

Asad, Asad L., and Matthew Clair. 2018. "Racialized Legal Status as a Social Determinant of Health." *Social Science & Medicine* 199: 19–28.

Bayley, David H. 1994. *Police for the Future.* Oxford: Oxford University Press.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2): 169–217.

Bell, Monica C. 2019. "Safety, Friendship, and Dreams." *Harvard Civil Rights-Civil Liberties Law Review* 54: 703–39.

Bell, Monica C. 2020a. "Anti-Segregation Policing." *NYU Law Review* 95: 650–765.

Bell, Monica C. 2020b. "Located Institutions: Neighborhood Frames, Residential Preferences, and the Case of Policing." *American Journal of Sociology* 125 (4): 917–73.

Blumstein, Alfred, and Joel Wallman. 2000. "The Recent Rise and Fall of American Violence." In *The Crime Drop in America*, edited by Alfred Blumstein and Joel Wallman, 1–12. Cambridge, UK: Cambridge University Press.

Boen, Courtney, Lisa Keister, and Brian Aronson. 2020. "Beyond Net Worth: Racial Differences in Wealth Portfolios and Black–White Health Inequality across the Life Course." *Journal of Health and Social Behavior* 61 (2): 153–69.

Boen, Courtney, Karen Kozlowski, and Karolyn Tyson. 2020. "'Toxic' Schools? How School Exposures during Adolescence Influence Trajectories of Health through Young Adulthood." *SSM-Population Health* 11.

Boylan, Richard T. 2019. "Imputation Methods Make Crime Studies Suspect: Detecting Biases via Regression Discontinuity." https://cear.gsu.edu/files/2019/05/Boylan-3.pdf?__cf_chl_jschl_tk__=pmd_xTxKsGlR4L7MQQOb3yoqbMviCif0SoYseqjGfL_cogc-1629908767-0-gqNtZGzNAeWjcnBszQhR.

Braga, Anthony A., Andrew V. Papachristos, and David M. Hureau. 2014. "The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis." *Justice Quarterly* 31 (4): 633–63.

Braga, Anthony A., and Glenn L. Pierce. 2005. "Disrupting Illegal Firearms Markets in Boston: The Effects of Operation Ceasefire on the Supple of New Handguns to Criminals." *Criminology & Public Policy* 4 (4): 717–48.

Brayne, Sarah. 2020. *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford: Oxford University Press.

Bronstein, Nathaniel. 2015. "Police Management and Quotas: Governance in the CompStat Era." *Columbia Journal of Law and Social Problems* 48: 543–81.

Bushway, Shawn, and Peter Reuter. 2008. "Economists' Contribution to the Study of Crime and the Criminal Justice System." *Crime and Justice* 37: 389–451.

Cancian, Francesca M. 1993. "Conflicts between Activist Research and Academic Success: Participatory Research and Alternative Strategies." *American Sociologist* 24: 92–106.

Chabria, Anita. 2017. "Fifty Men Commit Most Gun Crimes in Sacramento. Could Money, Mentoring Get Them to Stop?" *Sacramento Bee*, August 17. https://www.sacbee.com/news/local/article165689947.html.

Chalfin, Aaron, and Justin McCrary. 2017. "Criminal Deterrence: A Review of the Literature." *Journal of Economic Literature* 55 (1): 5–48.

Cihan, Abdullah, Yan Zhang, and Larry Hoover. 2012. "Police Response Time to In-Progress Burglary: A Multilevel Analysis." *Police Quarterly* 15 (3): 308–27.

Clayton, Abené. 2021. "This US City Was Working to Cut Its Police Budget in Half – Then Violent Crime Started to Rise." *Guardian*, March 19. https://www.theguardian.com/global-development/2021/mar/19/defund-the-police-gun-violence-surge-oakland.

Cohen, Lawrence H., Kathleen Cimbolic, Stephen R. Armeli, and Tanya R. Hettler. 1998. "Quantitative Assessment of Thriving." *Journal of Social Issues* 54 (2): 323–35.

Collins, Patricia Hill. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Coviello, Decio, and Nicola Persico. 2015. "An Economic Analysis of Black–White Disparities in the New York Police Department's Stop-and-Frisk Program." *Journal of Legal Studies* 44 (2): 315–60.

Delgado, Sheyla, Gina Moreno, Richard Espinobarros, and Jeffrey A. Butts. 2020. *Reported Crime in MAP Communities Compared with Other NYC Areas*. New York: John Jay College of Criminal Justice.

Desmond, Matthew, Andrew V. Papachristos, and David S. Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81 (5): 857–76.

Dixon, Ejeris. 2020. "Building Community Safety: Practical Steps toward Liberatory Transformation." In *Beyond Survival: Strategies and Stories from the Transformative Justice Movement*, edited by Ejeris Dixon and Leah Lakshmi Piepzna-Samarasinha, 15–25. Chico, CA: AK Press.

Donohue III, John J., and Steven D. Levitt. 2001. "The Impact of Race on Policing and Arrests." *Journal of Law and Economics* 44 (2): 367–94.

Doleac, Jennifer L., and Nicholas J. Sanders. 2015. Under the Cover of Darkness: How Ambient Light Influences Criminal Activity. *Review of Economics and Statistics* 97 (5):1093–1103.

Drakulich, Kevin, John Hagan, Devon Johnson, and Kevin H. Wozniak. 2017. "Race, Justice, Policing, and the 2016 American Presidential Election." *Du Bois Review* 14 (1): 7–33.

Du Bois, W.E. Burghardt. 1904. *Some Notes on Negro Crime, Particularly in Georgia*. Atlanta: Atlanta University Press.

Epp, Charles R., Steven Maynard-Moody, and Donald Haider-Markel. 2015. *Pulled Over: How Police Stops Define Race and Citizenship*. Chicago: University of Chicago Press.

Fagan, Jeffrey, and Elliott Ash. 2017. "New Policing, New Segregation: From Ferguson to New York." *Georgetown Law Journal Online* 106: 33–134.

Ferguson, Andrew Guthrie. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: NYU Press.

Fisk, Catherine L., and L. Song Richardson. 2017. "Police Unions." *George Washington Law Review* 85: 712–99.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University

Press.
Fryer, Roland G. Jr. 2019. “An Empirical Analysis of Racial Differences in Police Use of Force.” *Journal of Political Economy* 127 (3):
Fryer, Roland G. Jr. 2020. “What the Data Say about Police.” *Wall Street Journal*, June 22. https://www.wsj.com/articles/what-the-data-say-about-police-11592845959.
Garvin, Eugenia C., Carolyn C. Cannuscio, and Charles C. Branas. 2013. “Greening Vacant Lots to Reduce Violent Crime: A Randomised Controlled Trial.” *Injury Prevention* 19 (3): 198–203.
Gee, Gilbert C., and Chandra L. Ford. 2011. “Structural Racism and Health Inequities: Old Issues, New Directions.” *Du Bois Review* 8 (1): 115–32.
Geller, Amanda, Jeffrey Fagan, Tom Tyler, and Bruce G. Link. 2014. “Aggressive Policing and the Health of Young Urban Men.” *American Journal of Public Health* 104 (12): 2321–27.
Gerety, Rowan Moore. 2020. “An Alternative to Police that Police Can Get Behind.” *The Atlantic*, December 28. https://www.theatlantic.com/politics/archive/2020/12/cahoots-program-may-reduce-likelihood-of-police-violence/617477/.
Goff, Phillip Atiba, Tracey Lloyd, Amanda Geller, Steven Raphael and Jack Glaser. 2016. *The Science of Justice: Race, Arrests, and Police Use of Force.* New York: Center for Policing Equity.
Goncalves, Felipe, and Steven Mello. 2021. “A Few Bad Apples? Racial Bias in Policing.” *American Economic Review* 111 (5): 1406-41.
Gordon, Daanika. 2020. “The Police as Place-Consolidators: The Organizational Amplification of Urban Inequality.” *Law & Social Inquiry* 45 (1): 1–27.
Harcourt, Bernard E. 2007. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age.* Chicago: University of Chicago Press.
Harvard IQSS. 2021 “Our Flourishing Measure.” https://hfh.fas.harvard.edu/measuring-flourishing. (accessed May 21, 2021).
Hoekstra, Mark, and CarlyWill Sloan. 2020. “Does Race Matter for Police Use of Force? Evidence from 911 Calls.” NBER Working Paper 26774.
Hoggard, Corin. 2020. “Once Controversial, Advance Peace Gets City Funding to Tackle Fresno Violence.” ABC30, October 23. https://abc30.com/fresno-advance-peace-gun-violence-city-council/7254585/.
Jacobs, Jane. 1961 [2016]. *The Death and Life of Great American Cities.* New York: Vintage Books.
Jacques, Scott, and Richard Wright. 2015. *Code of the Suburb: Inside the World of Young Middle-Class Drug Dealers.* Chicago: University of Chicago Press.
Johnson, David J., Trevor Tress, Nicole Burkel, Carley Taylor, and Joseph Cesario. 2019. “Officer Characteristics and Racial Disparities in Fatal Officer-Involved Shootings.” *Proceedings of the National Academy of Sciences.* 116 (32): 15877–15882.
Johnson, David J., Trevor Tress, Nicole Burkel, Carley Taylor, and Joseph Cesario. 2020a. “Correction for Johnson et al., Officer Characteristics and Racial Disparities in Fatal Officer-Involved Shootings.” *Proceedings of the National Academy of Sciences.* 117 (16): 9127.
Johnson, David J., Trevor Tress, Nicole Burkel, Carley Taylor, and Joseph Cesario. 2020b. “Retraction.” *Proceedings of the National Academy of Sciences.* 117 (3): 18130.
Justice, Benjamin. 2021. “Hobbling: The Effects of Proactive Policing and Mass Imprisonment on Children’s Education.” *Annual Review of Law & Social Science* 17:3.1–3.21.
Keene, Kirsten, Kim Keating, and Pirkko Ahonen. 2016. *The Power of Stories: Enriching Program Research and Reporting.* Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Kerrison, Erin M., and Alyasah A. Sewell. 2020. “Negative Illness Feedbacks: High-Frisk Policing Reduces Civilian Reliance on ED Services.” *Health Services Research* 55 (52): 787–96.
Knox, Dean, and Jonathan Mummolo. 2020. “Making Inferences about Racial Disparities in Police Violence.” *Proceedings of the National Academy of Sciences* 117 (3): 1261–62.
Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. “Administrative Records Mask Racially Biased Policing.” *American Political Science Review* 114 (3): 619–37.
Kohler-Hausmann, Issa. 2019. “Eddie Murphy and the Dangers of Counterfactual Causal Thinking about Detecting Racial Discrimination.” *Northwestern Law Review* 113: 1163–1227.
Kurwa, Rahim. 2020. “Opposing and Policing Racial Integration: Evidence from the Housing Choice Voucher Program.” *Du Bois Review* 17 (2): 363–87.
Laniyonu, Ayobami. 2019. “The Political Consequences of Policing: Evidence from New York City.” *Political Behavior* 41: 527–58.

Legewie, Joscha, and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." *American Sociological Review* 84 (2): 220–47.

Levin, Benjamin. 2020. "What's Wrong with Police Unions?" *Columbia Law Review* 120 (5): 1333–1401.

Levitt, Steven D. 2004. "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not." *Journal of Economic Perspectives* 18 (1): 163–90.

Lu, Yi, Kaifeng Yang, and M. Blair Thomas. 2020. "Designing Performance Systems in Anticipation of Unintended Consequences: Experiences and Lessons from the Compstat-Based Performance Regime in NYPD." *Administration & Society* 53 (6): 907–36.

Ludwig, Jens, and John J. Donohue III. 2007. "More COPS." *Brookings Policy Brief*, March 1st. https://www.brookings.edu/research/more-cops/.

Lukachko, Alicia, Mark L. Hatzenbuehler, and Katherine M. Keyes. 2014. "Structural Racism and Myocardial Infarction in the United States." *Social Science & Medicine* 103: 42–50.

Lynch, James P., and John P. Jarvis. 2008. "Missing Data and Imputation in the Uniform Crime Reports and the Effects on National Estimates." *Journal of Contemporary Criminal Justice* 24 (1): 69–85.

Madden, Sidney, Sam Leeds, and Rodney Carmichael. 2020. "'I Want Us to Dream a Little Bigger': Noname and Mariame Kaba on Art And Abolition." *NPR,* December 19. https://www.npr.org/2020/12/19/948005131/i-want-us-to-dream-a-little-bigger-noname-and-mariame-kaba-on-art-and-abolition.

McCormack, Philip D., April Pattavina, and Paul E. Tracy. 2017. "Assessing the Coverage and Representativeness of the National Incident-Based Reporting System." *Crime & Delinquency* 63 (4): 493–516.

Meares, Tracey L. 2014. "The Law and Social Science of Stop and Frisk." *Annual Review of Law & Social Science* 10: 335–52.

Merolla, David M., and Omari Jackson. 2019. "Structural Racism as the Fundamental Cause of the Academic Achievement Gap." *Sociology Compass* 13 (6).

Milam, Adam J., Shani A Buggs, C. Debra M. Furr-Holden, Philip J. Leaf, Catherine P. Bradshaw, and Daniel Webster. 2013. "Changes in Attitudes toward Guns and Shootings Following Implementation of the Baltimore Safe Streets Intervention." *Journal of Urban Health* 93: 609–26.

Moskos, Peter. 2008. *Cop in the Hood: My Year Policing Baltimore's Eastern District.* Princeton: Princeton University Press.

O'Connell, Paul E. 2001. "Using Performance Data for Accountability: The New York City Police Department's CompStat Model of Police Management." In *Managing for Results 2002*, edited by Mark A. Abramson and John M. Kamensky. Lanham, 179–218. Lanham, MD: Rowman & Littlefield.

Okechukwu, Amaka. 2021. "Watching and Seeing: Recovering Abolitionist Possibilities in Black Community Practices of Safety and Security." *Du Bois Revew* 1–28.

One Million Experiments. 2021. One Million Experiments: A Virtual Zine Project c/o Project Nia & Interrupting Criminalization. https://millionexperiments.com/. (accessed April 7, 2021).

Pearl, Betsy. 2019. NeighborhoodStat: Strengthening Public Safety Through Community Empowerment. Center for American Progress, https://www.americanprogress.org/issues/criminal-justice/reports/2019/10/02/475220/neighborhoodstat-strengthening-public-safety-community-empowerment/.

Powell, John A. 2008. "Structural Racism: Building upon the Insights of John Calmore." *North Carolina Law Review* 86: 791–816.

Raskolnikov, Alex. 2020. "Criminal Deterrence: A Review of the Missing Literature." *Supreme Court Economic Review* 28: 1–59.

Reiss, Albert J. Jr. 1992. "Police Organization in the Twentieth Century." *Crime and Justice* 15: 51–97.

Roe, Emery. 1994. *Narrative Policy Analysis: Theory and Practice.* Durham: Duke University Press.

Roeder, Oliver, Lauren-Brooke Eisen, and Julia Bowling. 2015. *What Caused the Crime Decline?* New York: Brennan Center for Justice.

Rogers, Patricia J. 2008. "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions." *Evaluation* 14 (1): 29–48.

Rojek, Jeff, Richard Rosenfeld, and Scott Decker. 2012. "Policing Race: The Racial Stratification of Searches in Police Traffic Stops." *Criminology* 50 (4): 993–1024.

Ross, Cody T. 2015. "A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014." *PLOS One* 10 (11).

Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect.* Chicago: University of Chicago Press.

Sciacca, Annie. 2021. "Oakland Restores Police, Fire Services after Budget Cuts." *East Bay Times*, April

13. https://www.eastbaytimes.com/2021/04/13/oakland-restores-police-fire-services-after-budget-cuts/.

Sered, Danielle. 2019. *Until We Reckon: Violence, Mass Incarceration, and a Road to Repair.* New York: New Press.

Sewell, Abigail A. 2017. "Illness Associations of Police Violence: Differential Relationships by Ethnoracial Composition." *Sociological Forum* 32 (S1): S975–97.

Sewell, Alyasah Ali, Justin M. Feldman, Rashawn Ray, Keon L. Gilbert, Kevin A. Jefferson, and Hedwig Lee. 2021. "Illness Spillovers of Lethal Police Violence: The Significance of Gendered Marginalization." *Ethnic and Racial Studies* 44: 1089–1114.

Sharkey, Patrick. 2018. *Uneasy Peace: The Great Crime Decline, the Renewal of City Life, and the Next War on Violence.* New York: Norton.

Sharkey, Patrick, Gerard Torrats-Espinosa, and Delaram Takyar. 2017. "Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime." 82 (6): 1214–40.

Sierra-Arévalo, Michael. 2019. "The Commemoration of Death, Organizational Memory, and Police Culture." *Criminology* 57 (4): 632–58.

Simonson, Jocelyn. 2021. "Police Reform through a Power Lens." *Yale Law Journal* 130: 778–1049.

Slutkin, Gary, Charles Ransford, and R. Brent Decker. 2015. "Cure Violence: Treating Violence as a Contagious Disease." In *Envisioning Criminology*, edited by Michael D. Maltz and Stephen K. Rice, 43–56. Springer.

Small, Mario Luis. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life.* Oxford: Oxford University Press.

Small, Mario Luis. 2017. *Someone to Talk To.* Oxford: Oxford University Press.

Small, Mario L., and Devah Pager. 2020. "Sociological Perspectives on Racial Discrimination." *Journal of Economic Perspectives* 34 (2): 49–67.

Thomas, Amanda L. and Kevin T. Wolff. 2020. "Crime Distortion within the NYPD: A Potential Method for Estimating Crime Misclassification within CompStat Statistics." *Police Practice and Research* 22 (4): 1390–1407.

Van Wessel, Margit. 2018. "Narrative Assessment: A New Approach to Evaluation of Advocacy for Development." *Evaluation* 24 (4): 400–418.

VanderWeele, Tyler J. 2017. "On the Promotion of Human Flourishing." *PNAS* 114 (31): 8148–56.

VanderWeele, Tyler J. 2019. "Measures of Community Well-Being: A Template." *International Journal of Community Well-Being* 2: 253–75.

VanNatta, Michelle, and Mariame Kaba. 2013. "'We're In It for the Long Haul:' Alternatives to Incarceration for Youth in Conflict with the Law." https://chiyouthjustice.files.wordpress.com/2013/08/alt_to_incarcerationfinal3.pdf.

Vargas, Robert. 2020. "It's Time to Think Critically about the UChicago Crime Lab." *Chicago Maroon,* June 11. https://www.chicagomaroon.com/article/2020/6/11/time-think-critically-uchicago-crime-lab/.

Vidal, Jordi Blanes i., and Tom Kirchmaier. 2018. "The Effect of Police Response Time on Crime Clearance Rates." *Review of Economic Studies* 85 (2): 855–91.

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

Walker, Hannah L. 2020. "Targeted: The Mobilizing Effect of Perceptions of Unfair Policing Practices." *Journal of Politics* 82 (1): 119–34.

Warren, Mark R., José Calderón, Luke Aubry Kupscznk, Gregory Squires, and Celina Su. 2018. "Is Collaborative, Community-Engaged Scholarship More Rigorous than Traditional Scholarship? On Advocacy, Bias, and Social Science Research." *Urban Education* 53 (4): 445–72.

Webster, Daniel W., Jennifer Mendel Whitehill, Jon S. Vernick, and Frank C. Curriero. 2013. "Effects of Baltimore's *Safe Streets* Program on Gun Violence: A Replication of Chicago's *CeaseFire* Program." *Journal of Urban Health* 90 (1): 27–40.

Weisburd, David, Stephen D. Mastrofski, Ann Marie McNally, Rosann Greenspan, and James J. Willis. 2003. "Reforming to Preserve: CompStat and Strategic Problem Solving in American Policing." *Criminology & Public Policy* 2 (3): 421–56.

Whitehill, Jennifer M., Daniel W. Webster, Shannon Frattaroli, and Elizabeth M. Parker. 2014. "Interrupting Violence: How the CeaseFire Program Prevents Imminent Gun Violence through Conflict Mediation." *Journal of Urban Health* 91: 84–95.

Wood, George, Daria Roithmayr, and Andrew V. Papachristos. 2019. "The Network Structure of Police Misconduct." *Socius* 5: 1–18.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 49–70

The US Pretrial System: Balancing Individual Rights and Public Interests

Will Dobbie and Crystal S. Yang

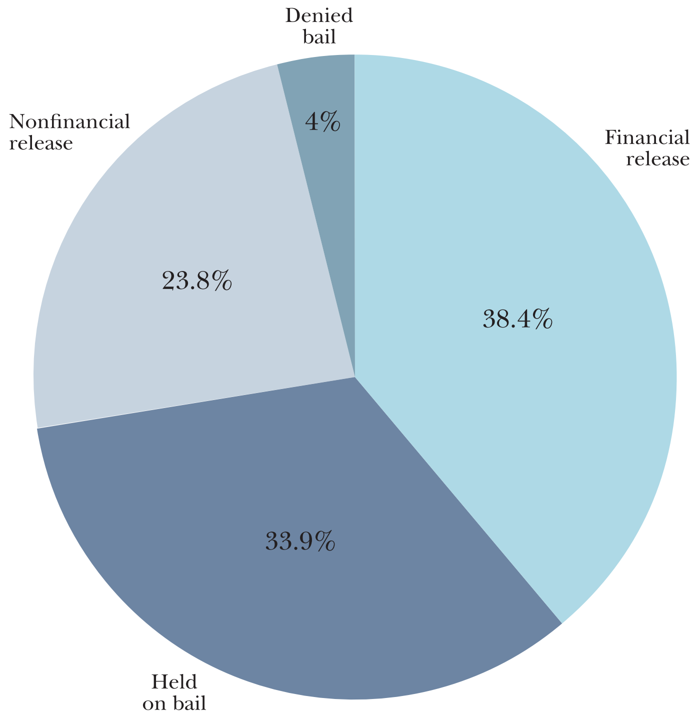
In most US jurisdictions, individuals accused of a crime appear for their first court appearance one to two days after being arrested by law enforcement. At this first appearance, sometimes referred to as a bail hearing or preliminary hearing, a judge is often tasked with determining a defendant's release or detention pending later adjudication—what is known as the "pretrial decision." On a yearly basis, decisions made at the pretrial stage affect roughly 11 million individuals who are arrested by law enforcement for an offense in the United States. In many parts of the country, these bail hearings last anywhere from ten seconds to a few minutes. Defendants are often video-conferenced in from the local jail while a judge briefly reviews the case and criminal history of the defendant, sometimes asking the defendant a few questions. If a prosecutor and/or defense attorney are present, they may also present their recommendations to the judge.

The judge in this bail or preliminary hearing has a range of options. For defendants who pose a minimal risk of flight or danger, the judge may simply release the defendant on the promise to return for future court proceedings and without any other conditions of release, sometimes known as "release on recognizance" or "personal recognizance." In some jurisdictions, judges also have the option of a "conditional release" with different types of non-monetary conditions, which can

■ *Will Dobbie is Professor of Public Policy at the Harvard Kennedy School and Crystal S. Yang is Professor of Law at the Harvard Law School, both in Cambridge, Massachusetts. Dobbie and Yang are Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are will_dobbie@hks.harvard.edu and cyang@law.harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.49.

Figure 1
Pretrial Outcomes for Felony Defendants



Source: State Court Processing Statistics, 2009 (US Department of Justice 1990–2009).
Note: Data do not include small share of individuals with missing pretrial outcomes or those released via emergency release. Because of rounding, numbers do not sum to 100 percent.

include regular reporting to a pretrial services officer, drug treatment or testing, orders that the defendant will have no contact with the alleged victim, or even electronic monitoring or home confinement. Another increasingly common option is the use of monetary or cash bail whereby defendants are offered release if they or someone on their behalf posts some amount of money, which is generally forfeited if they do not appear for trial or commit a new offense while out on release. Finally, some defendants, typically those who have been charged with the most serious crimes, are not offered pretrial release at all and are detained outright, sometimes known as "remand without bail."

Figure 1 documents the prevalence of this range of options in 2009 among a representative sample of felony defendants from the 75 largest US counties. Among these felony defendants, the overwhelming majority of defendants were assigned monetary bail (or financial conditions), with 34 percent of defendants held on bail because they did not pay the required amount to secure release and another 38 percent released on bail. Among the rest, 24 percent of defendants were released on non-financial conditions such as release on recognizance or conditional release, while another 4 percent were detained outright.

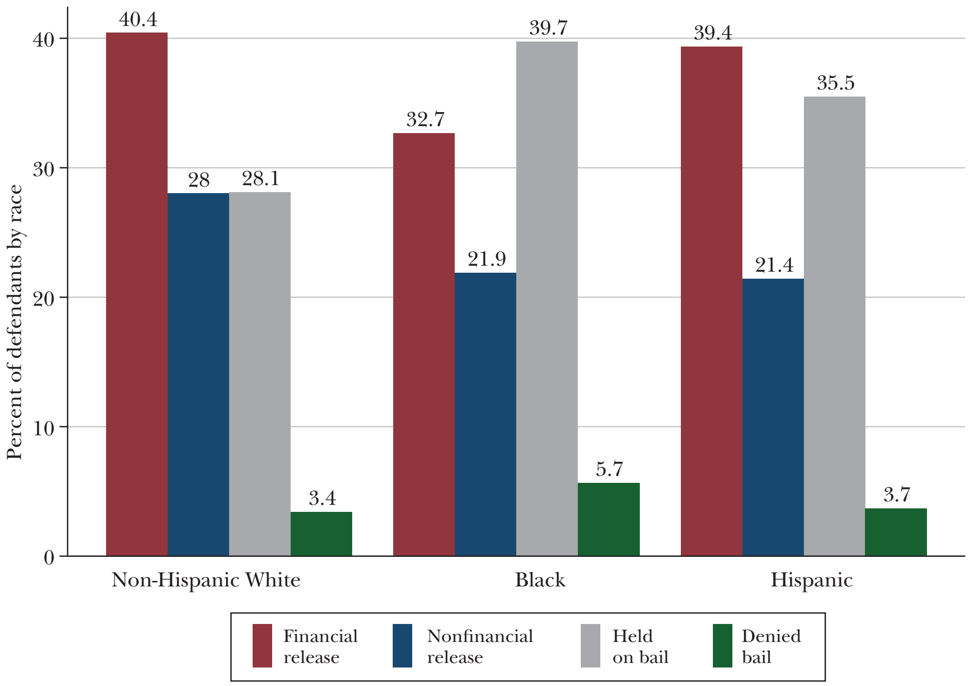
In making this pretrial decision, the judge is meant to balance several sets of concerns set out by law: the rights and liberty interests of the defendant, who after all has not yet been tried or convicted of any crime and thus is presumed to be innocent; having the defendant appear for a future trial and other required court appearances; and protecting society from additional crimes that the defendant might commit if released while awaiting trial. In addition to achieving this balance, the judge should seek to reach decisions that do not discriminate unlawfully on the basis of protected characteristics such as race, ethnicity, or gender.

In the next two sections of the paper, we discuss two important shifts in emphasis in the pretrial system in recent decades. The first is the shift toward having judges place a greater emphasis on public safety concerns or the risk of new crimes in their decision-making. The second is a shift toward increased use of monetary bail and subsequent decreased use of release without financial conditions. Due in part to these trends, the United States today detains roughly half a million individuals before trial at any given time, nearly twice as many as any other country in the world (Walmsley 2016). In per capita terms, the United States detains between two and 36 times as many individuals before trial as other high-income countries (Walmsley 2016). Indeed, today in some parts of the United States, over 75 percent of jail inmates (comprised of individuals awaiting trial or those serving relatively short sentences) are comprised of those detained pretrial.

These trends, as well as other existing patterns of the US pretrial system, raise two main concerns, which are the subjects of the following two sections of the paper. The first concern is that the extent of pretrial detention in the US generates costs to detainees that far outweigh the benefits to society. Within the criminal legal system, excessive pretrial detention is in tension with a foundational idea that people should be presumed innocent until proven guilty. Indeed, a sizeable number of detained individuals are eventually not found guilty of any offense. For example, among felony defendants charged in the 75 largest counties, one in five detained individuals later have their charges dismissed or are acquitted (Cohen and Reaves 2007). Even worse, the threat of pretrial detention increases the risk of wrongful conviction, by pressuring defendants to accept a plea bargain and get out of jail. In addition, pretrial detention may generate collateral consequences outside of the criminal legal system by disrupting defendants' lives, putting jobs, housing, and child custody at risk, among other harms. As a lawyer working as a public defender explained to the *New York Times* (reported in Pinto 2015):

> Our clients work in service-level positions where if you're gone for a day, you lose your job. People in need of caretaking—the elderly, the young—are left without caretakers. People who live in shelters, where if they miss their curfews, they lose their housing. Folks with immigration concerns are quicker to be put on the immigration radar. So when our clients have bail set, they suffer on the inside, they worry about what's happening on the outside, and when

Figure 2
Pretrial Outcomes for Felony Defendants, by Race/Ethnicity



Source: State Court Processing Statistics, 2009 (US Department of Justice 1990–2009).

> they get out, they come back to a world that's more difficult than the already difficult situation that they were in before.

A second concern is the presence of significant disparities in pretrial conditions and pretrial detention in most large US jurisdictions, contributing to the over-representation of low-income and minority individuals in the pretrial system. For example, rates of pretrial detention are significantly higher among Black and Hispanic individuals compared to non-Hispanic White individuals. Figure 2 presents pretrial outcomes among felony defendants in the largest 75 US counties, by race/ethnicity.

Among felony defendants arrested in 2009, 40 percent of Black individuals and 36 percent of Hispanic individuals were held on bail/financial conditions compared to 28 percent of non-Hispanic White individuals. In contrast, 22 percent of Black individuals and 21 percent of Hispanic individuals were released without financial conditions compared to 28 percent of non-Hispanic White individuals. Black and Hispanic individuals were also more likely to be denied bail and detained outright compared to non-Hispanic White individuals. These patterns mirror the all-too-common stories of Black and Hispanic individuals who, despite being first-time offenders accused of low-level crimes, spent months in pretrial detention with often devastating consequences.[1]

One plausible reason for these racial disparities is the extensive use of monetary bail. Critics of the current pretrial system argue that many jurisdictions set bail without adequate consideration of the defendant's ability to pay, and, as a result, that pretrial detention is determined by a defendant's wealth (which is correlated with race/ethnicity), not the likelihood of later appearing in court or the risk to the community during the pretrial period. These concerns have been long-standing. For example, at the signing of the federal Bail Reform Act of 1966, which sought to protect the right to pretrial release without the payment of money, President Lyndon Johnson remarked that "[b]ecause of the bail system, the scales of justice have been weighted for almost two centuries not with fact, nor law, nor mercy. They have been weighted with money." These issues led the Department of Justice (2016) to intervene in a recent case and conclude that the pretrial systems in many jurisdictions "are not only unconstitutional, but also constitute bad public policy."

Given these two major concerns with the pretrial system, researchers have sought to evaluate both the effectiveness and fairness of the system. However, empirical challenges have made it challenging to produce rigorous research on these topics. First, there is no general repository of data on pretrial decisions for the United States, so studies in this area focus on data from certain jurisdictions. Even among these studies, there are often no readily available datasets that include information on both pretrial decisions and long-term outcomes such as employment or the receipt of government assistance for a large number of individuals, making it difficult to assess the consequences of being detained pretrial. Second, an empirical analysis on these topics must confront difficult selection issues. Individuals who are detained before trial are not a random sample and are likely to be different from defendants who are not detained in a variety of ways not well-captured by existing data. These selection issues make it difficult to identify the causal effects of pretrial detention, which are required for evaluating whether the current pretrial system is appropriately balancing individual rights with societal benefits. These data and selection issues also make it challenging to evaluate the fairness of the current system, as defendants in different groups could be treated differently due to discrimination or due to legally relevant differences that are observed by the judge, but unobserved by the researcher.

In recent years, a growing empirical literature has made use of new data sources and quasi-experimental approaches to overcome these challenges and provide credible causal estimates of both the individual costs (such as loss of employment or government assistance) and public benefits (such as preventing non-appearance at court and new crimes) of cash bail and pretrial detention, and to do so in ways that

[1] As one vivid example, Gonnerman (2015) reports on the tragic death of Kalief Browder, who endured two years of solitary confinement while awaiting a trial that never happened.

illuminate the extent of racial discrimination in the pretrial system in a few jurisdictions with rich data. This burgeoning literature has produced convincing evidence that we should detain far fewer individuals before trial than we currently do and that the costs of pretrial detention are disproportionately borne by minority individuals due to forms of racial discrimination. We describe these studies below, highlighting areas worthy of future inquiry.

The closing sections of the paper discuss some policy implications of this recent work. We provide an overview of the wave of pretrial reform efforts happening across different cities and states, including greater use of non-monetary alternatives to bail. We also discuss the possibilities for judges to make greater use of algorithms or risk-assessments as a way of simultaneously reducing the extent of pretrial detention and pretrial crime, while also reducing racial disparities.

A Shift of Emphasis in the Goals of Pretrial Detention

Since the founding of the country, the principal objective of the US pretrial system has been to assure later appearance at court. Historically, future appearance at court was most often guaranteed through verbal pledges by others who assumed responsibility for having the accused appear for trial (the legal term is "sureties").

The Eighth Amendment to the US Constitution specifies that "[e]xcessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted." The Eighth Amendment protection meant that judges were supposed to release nearly all defendants before trial (except those charged with offenses subject to the death penalty), unless there was a serious flight risk. Again, the importance of pretrial release is grounded in the presumption of innocence, a fundamental right to protect defendants prior to any finding of guilt. This principle was embodied in the Judiciary Act of 1789, which stated that all non-capital defendants should be entitled to some form of bail. The Supreme Court has also stated that a defendant's bail cannot be set higher than an amount that is reasonably likely to ensure the defendant's presence at trial (*Stack v. Boyle*, 342 U.S. 1 [1951]).

In the past several decades, the pretrial system has shifted in its aims. Rather than focus exclusively on ensuring appearance for trial, the pretrial system today has also adopted an explicit aim of protecting the community from harm. Starting in the 1970s, in response to growing concerns about crime and public safety, jurisdictions began to authorize the detention of criminal defendants without bail if they were assessed to be dangerous to society—known as "preventive detention." For example, the federal Bail Reform Act of 1984 explicitly authorized judges to make bail determinations based on their assessment of each defendant's risk to the community. The 1984 Act states, among other things, that defendants should be granted bail "unless . . . such release will not reasonably assure the appearance of the person . . . or will endanger the safety of any other person or the community."

As a result, judges now make pretrial decisions based in part on their assessment of each defendant's risk of danger to the public. At present, the federal judicial system, along with at least 40 states, considers public safety explicitly as part of the release or detention decision. Only a few remaining jurisdictions, such as New York, base release decisions solely on an assessment of a defendant's risk of flight, although there have been numerous unsuccessful attempts to change New York law to include the criteria of public safety to a judge's pretrial decision-making. These competing objectives are embodied in the pretrial standards of the American Bar Association (2007), which states that the judicial decision of whether to release or detain a defendant requires judges to "strike an appropriate balance" between the competing societal interests of individual liberty, court appearance, and public safety. Similarly, the National Institute of Corrections states that "[t]he goal of bail setting is to maximize release while simultaneously maximizing court appearance and public safety" (Pilnik 2017). In other words, judges are theoretically required to make a proper tradeoff between a defendant's private liberty interests and the societal interests of court appearance and public safety, with neither set of goals being privileged over the other.

In practice, a non-negligible share of defendants does fail to appear in court (although very few abscond indefinitely) and some are rearrested for new offenses while out on release. For example, in 2009, among felony defendants in the 75 largest US counties, 17 percent of released individuals missed a court appearance and 3 percent were not returned to court within a year of release (Reaves 2013). Among these released individuals, 16 percent were rearrested for a new crime within a year of release, split roughly equally between felony and misdemeanor offenses (Reaves 2013).

While the objectives of the pretrial system are defined clearly by law, bail judges are granted substantial discretion in making assessments of flight risk and danger to the public. In many jurisdictions, denial of bail is often mandatory in first- or second-degree murder cases, but can also be imposed for other crimes, such as domestic violence, when the bail judge finds that no set of conditions for release will guarantee appearance or protect the community from the threat of harm posed by the defendant. In many jurisdictions, these bail judges may consider factors such as the nature of the alleged offense, the weight of the evidence against the defendant, any record of prior flight or bail violations, and the financial ability of the defendant to pay bail. As we will discuss later in the paper, some judges use a "risk score" based on these kinds of factors to offer guidance in the pretrial decision.

We hypothesize that the overall effect of considering public safety has likely been to increase rates of pretrial detention relative to a pretrial system that does not consider public safety. For instance, law enforcement often calls on judges to detain individuals who have been charged with violent offenses out of concern for safety (Barrett 2021). Anecdotally, many judges prioritize concerns about public safety, tipping the balance in marginal cases away from release to detention, either via monetary bail or outright detention. In instances where judges have released an individual who is later arrested for murder, for example, there is considerable public outcry and a demand for greater pretrial detention (McKinley 2017).

The Shifting Emphasis toward Monetary Bail

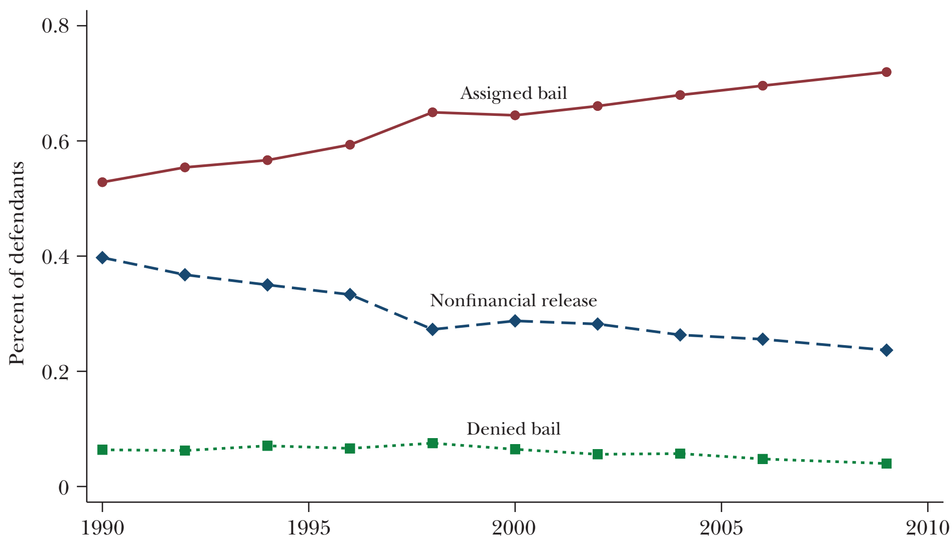
Another large contributor to the high rate of pretrial detention in the United States is the increasing use of monetary bail (also known as cash bail or bond) and the corresponding decreasing use of release on recognizance over the past several decades. Under the federal Bail Reform Act of 1966, the general presumption had been in favor of release without financial conditions out of a view that pretrial release should not be governed by a person's ability to pay. But in the last few decades of the twentieth century, the primary means of ensuring appearance and public safety in the United States has become the use of monetary bail, which in theory is meant to provide a financial incentive for defendants to refrain from engaging in pretrial misconduct.

Figure 3 documents these trends among felony defendants in the most populous 75 counties between 1990 and 2009. This figure shows that over this time period, felony defendants have been increasingly likely to be assigned monetary bail, from less than 60 percent in 1990 to 72 percent by 2009. In contrast, the share of felony defendants released without financial conditions has steadily decreased over this time period, falling from roughly 40 percent in 1990 to 24 percent by 2009.

Implementation of monetary bail varies across jurisdictions. In some places, defendants may need to post the full bail amount to secure release. In others, defendants are typically required to pay some fraction of the bail amount, such as 10 percent. Those who do not have the required deposit in cash can borrow from commercial bail bondsmen, who will often accept cars, houses, jewelry, or other forms of collateral, and who generally charge a nonrefundable fee, typically 10 percent of the bail amount, for their services. Another common type of monetary bail is an "unsecured" bond, which involves a promise by defendants to pay a certain amount of money if they do not return to court, but does not require an upfront payment to secure release. If the defendant fails to appear or commits a new crime (broadly known as "pretrial misconduct"), either the defendant or the bail bondsman is theoretically liable for the full value of the bail amount and forfeits any amount already paid. The amount of monetary bail may be determined by the judge or pre-specified in a "bail schedule," which determines bail amounts for each type or grade of offense, although a judge typically has discretion to change the recommended bail amount: for example, a bail schedule might specify that a Level 1 felony is associated with a $50,000 bail amount. Bail schedules are regularly used in California, Texas, and other states, although they have been criticized for failing to tailor amounts based on defendants' ability to pay.

One unique feature of the US bail system is that it is dominated by a $2 billion commercial bail bondsmen industry. According to the *Professional Bail Agents of the United States,* approximately 14,000 commercial bail agents nationwide secure the release of more than 2 million defendants annually (Cohen and Reaves 2007). As described above, bail bondsmen are permitted to post bail in exchange for nonrefundable payments from the defendants and the promise that they would find the defendants and return them if they failed to appear. While some argue that

Figure 3
Trends in Pretrial Outcomes among Felony Defendants, 1990–2009



Source: State Court Processing Statistics, 1990–2009 (US Department of Justice 1990–2009).

the industry is highly effective at ensuring that defendants appear at court, other recent accounts have documented purported anti-competitive practices, with the bail industry accused of "collud[ing] with lawyers, the police, jail officials and even judges to make sure that bail is high and that attractive clients are funneled to them" (Liptak 2008). The commercial bail bondsmen industry is illegal in nearly all other countries and exists today in only the United States and the Philippines.

The growing use of monetary bail in many jurisdictions has resulted directly in high pretrial detention rates, as many defendants are unable or unwilling to pay even relatively small monetary bail amounts. A typical sum of monetary bail imposed by judges is $10,000 (Reaves 2013), but 40 percent of Americans are unable to pay an "unexpected expense" of $400 (Board of Governors of the Federal Reserve 2018). In New York City, an estimated 46 percent of all non-felony defendants and 30 percent of all felony defendants were detained before trial in 2013 because they were unable or unwilling to post bail set at $500 or less (New York City Criminal Justice Agency 2014). This is not surprising once one considers that the typical defendant is quite poor. For example, among individuals detained in Philadelphia and Miami between 2006 to 2014, only 32.0 percent were formally employed in the year prior to arrest and the average annual formal-sector income is $4,524 (Dobbie, Goldin, and Yang 2018).

In addition, monetary bail may not serve its intended deterrent function very well. Using a natural experiment in Philadelphia, Ouss and Stevenson (2021) find that a reform which led to a sharp decrease in the use of monetary bail without a change to the overall pretrial release rate did not significantly increase failure to appear or new criminal activity rates among those who were released.

Evidence on Tradeoffs of Pretrial Detention

How can researchers evaluate whether bail judges are properly balancing the private interests of defendants (costs of pretrial detention) versus the societal interests of ensuring court appearance and public safety (benefits of pretrial detention)?

When thinking about how to estimate the magnitude of the cost and benefits of pretrial decisions, a hypothetical social experiment might randomly offer pretrial release to a wide range of offenders and then observe the results. Of course, such an experiment is impractical for many reasons, including the fact that explicitly randomized detention would outrage the principles of justice. Because such randomization is unethical and infeasible, any simple comparisons of the outcomes of released and detained individuals almost certainly suffer from selection bias, given that judges may be less likely to release an individual at higher risk of flight or danger. In addition, empirical research in this area is also complicated by the fact that, until recently, few datasets linked pretrial decisions to defendant and societal outcomes, both in and out of the criminal legal system.

However, recent data and methodological advances have allowed researchers to identify causal estimates of some key costs and benefits of pretrial detention. In particular, researchers have been able to obtain readily available court data (often in jurisdictions with permissive public records access laws) that also contain individual identifiers, which in turn can be linked to administrative data on key outcomes. The results make a strong argument that the costs of pretrial detention (and cash bail) almost certainly far outweigh the social benefits, at least for the "marginal" defendants for whom judges disagree about whether to release or detain.

The key methodological insight behind this recent literature was to observe that in some jurisdictions, defendants are assigned more-or-less randomly to judges for their pretrial hearing and that judges differ systematically on decisions about bail conditions. Indeed, given the discretion that bail judges have, there is often substantial variability in pretrial decisions, even among judges assigned similar defendants in the same court system (Yang 2017). Why this variability exists is not well known, but judges may differ in how they trade off the private interests of the accused versus societal interests or in how they assess the risks of pretrial misconduct.

Whatever the reason, as a result of this variability across judges, some defendants are more likely to be assigned a relatively low bail amount or release on their own recognizance (resulting in pretrial release) while other defendants with the same characteristics are more likely to be assigned a higher monetary bail amount or no bail at all (resulting in pretrial detention), based only on whether they were randomly assigned to a "lenient" or a "strict" judge. This naturally occurring source of variation approximates the hypothetical ideal experiment. In addition, bail judges are different from the trial and sentencing judges in many jurisdictions who are assigned through a different process. This institutional feature allows researchers to go one step further and to identify the causal effect of being assigned to a lenient bail judge as opposed to a lenient trial or sentencing judge.

Building on these features, a new literature has utilized the so-called "judge instrumental-variable" empirical design to recover the causal effect of pretrial detention for individuals at the margin of detention.[2] Some recent papers using this approach include Gupta, Hansman, and Frenchman (2016) using data from courts in Philadelphia and Pittsburgh, Leslie and Pope (2017) using data from New York City, Dobbie, Goldin, and Yang (2018) using data from counties that contain Philadelphia and Miami, Stevenson (2018) using data from Philadelphia, and Didwania (2020) using data from federal district courts. Each of these papers also leverages the richness of court-specific data to measure the effects of pretrial detention on case outcomes and when available, pretrial flight and pre- and post-trial crime. In addition, Dobbie, Goldin, and Yang (2018) link these court-specific data to administrative tax records at the Internal Revenue Service to examine the effects of pretrial detention on post-trial economic outcomes such as formal-sector employment and the take-up of government benefits.

These papers focus on several primary sets of key outcomes from pretrial detention, including the likelihood of being found guilty, the likelihood of appearing in court, the likelihood of being arrested for a new crime, and later economic outcomes such as formal-sector employment and the take-up of government benefits. We discuss each of these outcomes in turn.

Across all these papers, the authors find that being detained pretrial adversely affects a defendant's case. In Dobbie, Goldin, and Yang (2018), for example, the authors use data on over 400,000 criminal defendants arrested in Miami and Philadelphia from 2007 to 2014. They find that being detained before trial increases the probability of being found guilty by 14 percentage points—a 24 percent change from the mean for released defendants. The increase in convictions is largely driven by a higher probability of pleading guilty. The authors interpret these results as suggesting that pretrial detention primarily affects case outcomes by weakening defendants' bargaining positions before trial. Dobbie, Goldin, and Yang (2018) also show that pretrial detention has a small and statistically insignificant effect on *post-trial* incarceration, which is consistent with the story that defendants are pleading guilty to "time served"—that is, a defendant who has been held in jail before trial and pleads guilty can go home immediately with no further incarceration. The

[2] Formally, the conditions necessary to interpret these judge instrumental-variable estimates as the causal impact of pretrial detention for individuals at the margin of detention are: 1) there is a first-stage relationship between judge assignment and the probability of pretrial detention; 2) judge assignment only impacts defendant outcomes through the probability of being detained; and 3) any defendant released by a strict judge would also be released by a more lenient judge, and any defendant detained by a lenient judge would also be detained by a more strict judge. We direct the reader to Dobbie, Goldin, and Yang (2018) and Frandsen, Lefgren, and Leslie (2019) for a discussion of the extent to which these conditions are likely to hold in the pretrial context.

other studies mentioned above find qualitatively similar effects. Gupta, Hansman, and Frenchman (2016) examine the effects of being assigned monetary bail, rather than being detained directly, finding that monetary bail increases the likelihood of being convicted in a sample of cases from Philadelphia and Pittsburgh.

Substantial anecdotal evidence suggests that even a short stint of pretrial detention can have significant costs for defendants outside the criminal legal system: for example, they may be fired from a job or lose their space in a housing shelter if they are detained for several days. Dobbie, Goldin, and Yang (2018) study one set of medium-term effects of pretrial detention by linking defendants to later economic outcomes as measured in administrative tax records in their sample of Miami and Philadelphia cases. In this study, pretrial detention decreases the probability of employment in the formal labor market three to four years after the bail hearing by 9.4 percentage points in their data (a 25 percent decrease from the released defendant mean). Pretrial detention also decreases the probability that the defendant takes up unemployment insurance benefits within three to four years after case disposition and decreases the take-up of Earned Income Tax Credit (EITC) benefits over the same time period. The authors interpret these results as the stigma of a criminal conviction lowering defendants' prospects in the formal labor market (as discussed in Pager 2003; Agan and Starr 2018), which in turn limits their eligibility for employment-related benefits.

Recall that a judge will seek to balance the individual presumption of innocence with the societal benefits of preventing pretrial flight and crime. How well is the current pretrial system doing in terms of achieving these societal benefits?

To evaluate the risk of additional criminal activity, the crime effect of pretrial detention can be split into pretrial and post-trial crime. A defendant who is detained before case disposition will, because of the "incapacitation" effect, be by definition unable to commit new crimes in the community or fail to appear in court. However, if post-trial crime rates are increased by pretrial detention, then the net effect of pretrial detention on all future crime (combining pre- and post-trial crime) is ambiguous.

For example, in their study of data from Philadelphia and Miami-Dade, Dobbie, Goldin, and Yang (2018) find that pretrial detention decreases the probability of failing to appear in court by 15.6 percentage points. The effect on future crime is driven by offsetting short-run incapacitation and medium-term criminogenic effects—that is, detention causes the likelihood of re-arrest before case disposition to fall by 18.9 percentage points and the likelihood of re-arrest following case disposition to *increase* by 12.1 percentage points.[3] The authors argue that this criminogenic effect of pretrial detention may be due to decreased attachment to the formal labor market described previously—that is, those who lose their jobs may be more likely to commit new crimes. Leslie and Pope (2017) similarly find partially offsetting effects in New York City felony cases. In contrast, Gupta, Hansman, and Frenchman (2016)

[3] One important qualification of these findings is that it is challenging to identify new criminal activity from new re-arrest, which is an imperfect proxy for criminal activity.

find that being assigned monetary bail in Philadelphia and Pittsburgh has only a negligible effect on failure to appear in court, but leads to a 0.7 percentage point yearly increase in the probability of committing a new crime. This evidence suggests that when analyzing the effects of pretrial detention on crime, both pretrial and post-trial criminal activity should be considered.

So how can we combine these estimates to evaluate how the pretrial system is doing in trading off a defendant's private interests against societal benefits? Any attempt to weigh the costs and benefits of detentions is necessarily a rough back-of-the-envelope calculation. However, Dobbie, Goldin, and Yang (2018) conduct a partial calculation that considers the administrative costs of jail, the costs of apprehending individuals who fail to appear, the costs of future criminality (both pre- and post-trial), and the labor market impacts on defendants. By their estimate, the net social cost of pretrial detention for the typical marginal defendant is between \$55,143 and \$99,124. The costs to defendants with no recent criminal history will be especially high, given the significant collateral consequences of having a criminal conviction on labor market outcomes, the offsetting criminogenic effect of pretrial detention in the medium-run, and the relatively low costs of apprehending defendants who fail to appear in court.

This estimate should be interpreted with caution. On one side, it does not include possible benefits of general deterrence effects: that is, the possibility that detaining individuals before trial reduces crime among those who are not arrested in the first place. On the other side, it does not include potential non-economic costs to defendants who are detained—like inability to provide care to family members and loss of freedom. However, the estimates suggest that the current pretrial system imposes substantial short- and long-term economic harms on detained defendants that are likely not justified by the societal benefits of reducing the risk of new crime and non-appearance at court.

This conclusion is reinforced by recent work showing that low-cost text message reminders can dramatically reduce failure to appear rates, suggesting that a large share of defendants who fail to appear are not intentionally skipping court but are effectively unaware of court dates (Fishbane, Ouss, and Shah 2020). Identifying ways to reduce the risks of new criminal activity or failure to appear without detention is a promising area for future research.

While the recent literature has made progress in documenting some of the main costs and benefits of pre-trial detention, more remains to be done. One important caveat to recent work in this area is that all of the estimates from the judge instrumental-variable approach are conceptually based on defendants at the margin of release or defendants for whom judges disagree on whether to release or detain, not the average defendant who might be released. Dobbie, Goldin, and Yang (2018) show that only about 13 percent of individuals in their sample are at the margin of release, for example, with these individuals being much more likely to be charged with misdemeanors and nonviolent offenses. Thus, the calculations may under- or overestimate the benefits of much larger changes to the pretrial system, such as completely eliminating cash bail as many jurisdictions are considering or doing, and releasing nearly all defendants before trial (including defendants charged with felonies and violent offenses). A fruitful area for future research is to explore the costs and benefits of pretrial detention for these *inframarginal* defendants. Another important step for future research is to expand the range of outcomes studied to evaluate other costs of pretrial detention like effects on mental and physical health of defendants and on outcomes for the dependents and families of defendants affected by the pretrial system.

Evidence on Racial Discrimination in Pretrial Decisions

Most large US jurisdictions exhibit significant racial and ethnic disparities in the imposition of pretrial conditions and pretrial detention. Nationally representative data on felony defendants in state courts show that on average, Black and Hispanic defendants are substantially more likely to be detained before trial compared to non-Hispanic White defendants, even after controlling for observable and legally relevant charge and defendant characteristics (Demuth 2003; Baradaran and McIntyre 2013). In addition, Black and Hispanic defendants are generally more likely to be assigned monetary bail and higher monetary bail amounts, compared to observably similar non-Hispanic White defendants (Demuth 2003; Demuth and Steffensmeier 2004; Schlesinger 2005).[4]

While these past studies reveal significant racial and ethnic disparities in the pretrial system, this conceptual and methodological approach of controlling for an array of observable variables and then looking at the coefficient on an indicator variable for race/ethnicity may not isolate discrimination in pretrial decisions. One concern is that certain legally relevant differences in pretrial misconduct risk may be observed by judges, but not by the econometrician, and this omitted variables bias can account for some observed racial and ethnic disparities in pretrial detention. For example, suppose that a judge appropriately considers a defendants' employment status or ties to the community when making a pretrial decision, and these factors differ by race. If controls for these variables are not included, it could make the coefficient on race look larger than it would otherwise be.

On the other hand, controlling for observable case and defendant characteristics can also hide the way in which discrimination can operate through seemingly race-neutral characteristics. For example, consider a hypothetical example in which White defendants are more likely than Black defendants to be arrested for offenses involving powdered versus crack cocaine. Suppose further that judges release a higher proportion of defendants arrested for offenses involving powder cocaine versus crack cocaine, but are "race-blind" among individuals charged with the same type of offense. Conditioning on the specific offense, a researcher would find no disparity in release rates. Yet this "race-blind" rule with respect to powder cocaine

[4] We focus on the fair treatment of individuals of different races or ethnicities, directing the reader to Yang and Dobbie (2019) for a discussion of other forms of unfairness in the pretrial system.

versus crack cocaine may well be due to unjustified discrimination. These concerns suggest considerable caution with an approach that conditions on as many observables as possible to "explain away" observed racial disparities. Some scholars call this form of bias "included variable bias"—the idea that controlling for non-racial variables could mask the existence of unwarranted discrimination (Ayres 2010).

To shed light on various forms of discrimination that may exist in the pretrial system, we think that economists should better adapt their methodological approaches to leading sociological, psychological, and/or legal definitions of discrimination. Of particular importance are the two main legal doctrines of discrimination in the United States: disparate treatment doctrine and disparate impact doctrine.

The disparate treatment doctrine prohibits policies or practices motivated by a "discriminatory purpose" and thus requires proof of intent.[5] In the pretrial context, there is disparate treatment if, for example, a bail judge intentionally sets higher monetary bail amounts for Black versus White defendants because of explicit bias against Black individuals. Importantly, a court is far less likely to find disparate treatment if racial disparities are unconscious, as could occur if a bail judge sets higher monetary bail amounts for Black versus White defendants because of implicit bias or unconscious stereotypes. For economists, it is difficult to identify a statistical test that conclusively tests for purposeful disparate treatment. One can compare the treatment of different groups in various ways that point to disparities, such as by controlling for observables or using an audit study. But to meet the legal disparate treatment standard, one would need to augment this statistical comparison with non-statistical evidence showing or strongly suggesting discriminatory intent on the part of judges, which can be difficult to establish.

However, existing models of racial discrimination in the economics literature often envision decision-makers who make conscious and intentional decisions based on race. In particular, the two canonical economic models of discrimination are 1) statistical discrimination whereby a person accurately uses race to make a prediction about unobserved potential outcomes, such as pretrial misconduct (for example, Aigner and Cain 1977) and 2) Becker taste-based discrimination whereby a person intentionally sets different decision thresholds for different racial groups, such as different thresholds for pretrial release (for example, Becker 1957, 1993). To distinguish between these approaches empirically, Becker (1957, 1993) proposed an "outcome test" that compares the success or failure of decisions across groups at the margin. In the context of pretrial decisions, the idea is that marginal

[5] The disparate treatment doctrine comes from the Equal Protection Clause of the US Constitution's Fourteenth Amendment. It was formalized in the landmark case *Washington v. Davis* (426 US 229 [1976]), where the Supreme Court explained that the "basic equal protection principle that the invidious quality of a law claimed to be racially discriminatory must ultimately be traced to a racially discriminatory purpose." Later, in *McCleskey v. Kemp* (481 US 279 [1987]), the Court rejected a challenge to Georgia's capital punishment scheme—despite statistical evidence showing large racial disparities in death penalty rates—because the evidence was "clearly insufficient to support an inference that any of the decisionmakers in [the defendant's] case acted with discriminatory purpose."

White defendants will have higher rates of pretrial misconduct than marginal Black defendants if bail judges are racially biased against Black defendants. Recent work has clarified that outcomes can vary across groups at the margin due to Becker taste-based discrimination, inaccurate racial stereotypes (such as those modeled by Bordalo et al. 2016), and potentially other form of racial bias (Arnold, Dobbie, and Yang 2018; Hull 2021; Gelbach 2021). The outcome test also captures de facto racial bias that might arise though seemingly race-neutral characteristics such as type of crime and neighborhood by allowing non-race characteristics to differ for marginal White and marginal Black defendants.[6]

The outcome test has been difficult to implement in practice because the concept relies on marginal comparisons, while the available data is typically on averages. For example, if White and Black defendants have different risk distributions for pretrial misconduct—the well-known "inframarginality problem (for example, Ayres 2002)—then inferring the required marginal misconduct rates from the average misconduct rates is problematic.

However, the "judge instrumental-variable approach" (described earlier) actually provides causal estimates for individuals at the margin of release—thus allowing researchers to measure the misconduct rates of White and Black defendants at the margin of release and providing a test of whether judges are setting different decision thresholds for White and Black defendants, whether that be due Becker taste-based discrimination, inaccurate stereotypes, or some other form of racial bias.[7] Arnold, Dobbie, and Yang (2018) use this approach with data from Miami-Dade and Philadelphia to show that marginally released White defendants are much more likely to be re-arrested compared to marginally released Black defendants. Their results suggest that judges make substantial errors in predicting in pretrial risk in a way that exaggerates the dangerousness of Black defendants—a form of "racial bias" distinct from Becker taste-based discrimination that originally motivated such outcome tests.

The results from Arnold, Dobbie, and Yang (2018), along with related discussion in Hull (2021), also make clear that outcome tests are purposely narrow, isolating racial bias from other important sources of disparities such as illegal statistical discrimination on the basis of race. The fact that outcome tests also combine behavior that is clearly disparate treatment, such as Becker taste-based discrimination, and behavior that a court may not find illegal under this standard, such as unconscious/implicit bias or stereotyping, further complicates the interpretation and use of these tests.

[6] The outcome test cannot be used to test for more restrictive definitions of racial bias that hold fixed all non-race characteristics of White and Black defendants without additional restrictions: for example, see the discussions in Canay, Mogstad, and Mountjoy (2020) and Hull (2021).

[7] New work by Grau and Vergara (2020) proposes an alternative implementation of the outcome test that relies on using the predicted status (or propensity scores) of defendants to identify defendants at the margin of release. Using this novel approach, the authors find evidence of racial bias among minority defendants in the pretrial system in Chile.

These conceptual problems with the outcome test and the disparate treatment doctrine more generally lead us to the disparate impact doctrine of discrimination. Under this doctrine, a policy or practice is discriminatory if it leads to an adverse impact on a protected class (say, by race or ethnicity) and the decision-maker cannot offer a substantial legitimate justification for the adverse impact.[8] Under existing law, disparate impact only applies in certain contexts, because it stems from statutory rules rather than constitutional law. As for the pretrial context, one potential avenue for bringing a disparate impact claim could be through Title VI of the 1964 Civil Rights Act, which covers all programs and activities receiving federal financial assistance and generally includes the state and local courts considered in our analysis of the pretrial setting.[9] However, we are not aware of such a case to date.

Regardless of whether a disparate impact case would have legal merit in the pretrial context, we view the doctrine as shedding light on pervasive forms of disparity and discrimination that we believe deserve more attention in the economics literature. Other fields such as sociology have long recognized that discrimination "may or may not result from prejudice or animus and may or may not be intentional in nature" (in this journal, Small and Pager 2020). Pretrial decision-making not intentionally premised on an individual's race can still produce unjustified disparities in society.

Recall that the only legally allowable justification to set more stringent conditions of pretrial release (such as cash bail) or to detain outright individuals before trial is the risk of pretrial misconduct (whether in the form of not appearing at trial or committing a crime). Thus, the disparate impact standard is violated if a judge sets higher monetary bail amounts for Black versus White defendants with the same potential for pretrial misconduct for any reason: "direct discrimination" arising from the consideration of race; unconscious or implicit discrimination; or "indirect discrimination" coming from the consideration of non-race characteristics that nevertheless lead to an adverse impact on a racial group. The disparate impact doctrine similarly prohibits any unjustified differences in treatment based on other protected characteristics like ethnicity or gender.

Unfortunately, much of the existing work in economics does not accommodate or consider discrimination under the disparate impact doctrine. The ideal statistical test for disparate impact would compare the treatment of individuals with identical potential outcomes—which is conceptually distinct from comparing the treatment

[8] The disparate impact doctrine was formalized in the landmark case of *Griggs v. Duke Power Co.* (401 US 424 [1971]). In that case, the US Supreme Court found that the Duke Power Company's policy of requiring that all employees have a high school diploma to be considered for promotion was illegal disparate impact because the requirement had little to no relationship to job performance (the legitimate justification in this case) and drastically limited the eligibility of Black employees.

[9] In *United States v. Maricopa County* (915 F. Supp. 2d 1073 [D. Ariz. 2012]), for example, the US Department of Justice filed suit under Title VI of the Civil Rights Act of 1964, alleging that the Maricopa County Sheriff's Office and Maricopa County were in violation of the disparate impact standard due to the failure "to develop and implement policies and practices to ensure LEP [limited English proficient] Latino inmates have equal access to jail services."

of individual with identical non-race characteristics. In the context of pretrial decisions, we would like to compare the pretrial decisions of White and Black defendants with identical potential for pretrial misconduct, without any additional controls for observable characteristics such as alleged crime type or criminal history.

Algorithms and Judges: Less Detention, Less Pretrial Misconduct, and Reduced Racial Disparities?

The current pretrial system seems to be performing poorly both in terms of balancing tradeoffs and treating individuals fairly by race. Is it possible to develop a set of guidelines or decision aids for bail judges that could improve on the current pretrial system in all dimensions: that is, simultaneously reduce pretrial detention, reduce crime committed by those who are released, and have an improved likelihood of defendants appearing at trial—all while reducing racial disparities?

The possibility of such an outcome is greatest if there are pervasive errors in pretrial decision, in the sense that judges are detaining excessive numbers of low-risk defendants and releasing excessive numbers of high-risk individuals, in a way that might be corrected by the use of guidelines or other decision-making aids.

Several studies have found evidence of such pervasive errors, including Arnold, Dobbie, and Yang (2018), Kleinberg et al. (2018), and Arnold, Dobbie, and Hull (2020). For example, Kleinberg et al. (2018) find evidence of pervasive errors in pretrial decisions for bail judges in New York City. They find that pretrial decisions of these judges are characterized by internal inconsistency and that judges struggle the most with high-risk cases. Not only do judges differ considerably from each other in their pretrial decisions, but even within the *same* judge, decisions for similar defendants vary quite a bit. The authors construct an algorithm to predict pretrial misconduct risk and find that if pretrial decisions were made using the algorithm (which does not include race as an input in prediction), the algorithm could be used to detain the same number of people with less pretrial crime, or detain fewer people with the same pretrial crime, or choose some mixture of reducing both pretrial crime and pretrial detention rates—all while simultaneously reducing racial disparities. In particular, the gains on racial equity are seen throughout the distribution of judges.

Such algorithms or “risk assessment” tools are already used quite widely in the pretrial system. Perhaps the best-known of the pretrial risk assessment tools is the Arnold Ventures Public Safety Assessment (PSA), which is now in use in over 39 jurisdictions around the country. The PSA uses nine factors to assess the likelihood of pretrial success: 1) age at current arrest; 2) current violent offense (in some cases whether the violent offense occurred when the defendant was 20 years old or younger); 3) pending charge at the time of the arrest; 4) prior misdemeanor conviction; 5) prior felony conviction (or in some cases any prior conviction, misdemeanor or felony); 6) prior violent conviction; 7) prior failure to appear in the past two years; 8) prior failure to appear older than two years; and 9) prior sentence to incarceration. These factors are weighted and then used to predict three outcomes:

failure to appear pretrial, new criminal activity while on pretrial release, and new violent criminal activity while on pretrial release. The PSA scores are usually accompanied by a Decision Making Framework that includes a matrix matching each combination of PSA scores and charges to a recommended pretrial decision for judges.

Studying the impact of such risk assessment tools on pretrial decision-making (both in terms of balancing tradeoffs and fairness) is an important area for future work. After all, as a practical matter, pretrial decisions are unlikely to be made solely on the basis of algorithmic predictions. Indeed, to the extent that judges often override the recommendations of algorithms, as has been shown in the sentencing context (Doleac and Stevenson 2019), it may be much more difficult to achieve improvements. Greiner et al. (2020) provide a detailed interim case study of how the Arnold Ventures PSA was implemented in Dane County, Wisconsin, and evalute the impact of the PSA using a randomized control trial. The early-stage results found that providing judges with these recommendations somewhat changed pretrial decisions, but had no statistically significant effect on days of pretrial incarceration, failure to appear rates, or new criminal activity rates.

Of course, algorithms are not a panacea for the pretrial system. Indeed, a robust literature has documented a variety of issues that algorithms can introduce ranging from bias being "baked in" to the algorithm, to lack of transparency and accountability. But this is an area where economists have much to contribute in thinking about how to design risk assessment tools that are efficient and equitable.

Conclusion

The US pretrial system seeks to balance the individual rights of defendants against the societal goals of ensuring court appearances and public safety—and to achieve these tradeoffs as equitably as possible. However, the research literature in this area raises doubts about the performance of the system. Pretrial detention imposes large economic and personal costs, due to the significant collateral consequences of having a criminal conviction on labor market outcomes and large social costs stemming from the criminogenic effects of pretrial detention. In addition, there are relatively small benefits of pretrial detention, due to the low costs of apprehending defendants who fail to appear in court and the relatively low-level crimes that occur when individuals are released before trial. Taking a range of costs and benefits into account, the existing evidence suggests that we should detain far fewer individuals before trial than we currently do. Moreover, the costs of cash bail and pretrial detention are disproportionately borne by Black and Hispanic individuals, giving rise to large racial differences in pretrial detention that cannot be explained by differences in pretrial misconduct risk.

Looking ahead, we highlight three sets of developments that bear particular attention in the near- and medium-term, and which we believe should guide the direction of future economics research in this area.

First, a wave of support seems to be gathering for reform of pretrial systems in many jurisdictions. In a national survey of registered voters in 2018, 72 percent supported limiting the length of pretrial detention, and over 70 percent were in favor of providing pretrial support services for those with addiction or mental health issues (Pretrial Justice Institute 2018). Some jurisdictions, like New Mexico, New Jersey, and California, have effectively eliminated cash bail, with the hope of significantly decreasing pretrial detention rates among low-risk defendants. In addition, a wave of community-based efforts to change the current pretrial system has also been spreading, with charitable organizations like the Bronx Freedom Fund and the Brooklyn Community Bail posting bail for individuals held on misdemeanor charges when bail is set at $2,000 or less. Future research should study these changes to the pretrial system. In doing so, researchers should also expand the list of potentially relevant outcomes to study to include health and family outcomes for individuals interacting with the pretrial system.

Second, there remains much need for further conceptual and empirical work that tests for discrimination in the pretrial system, given the large and pervasive racial disparities in detention rates. As other disciplines have long recognized, discrimination can result even when a decision-maker is not acting consciously and intentionally on the basis of race. We urge economists to think more broadly about the ways in which discrimination can manifest in the pretrial system and we view the legal doctrines of disparate treatment and disparate impact as important guideposts.

Finally, many jurisdictions are providing judges with new information and tools with the hope of improving the efficiency and fairness of pretrial decisions. The most prominent example, the Arnold Ventures Public Safety Assessment, was mentioned earlier. Other jurisdictions are piloting programs with other kinds of individualized feedback and behavioral interventions available to judges. These field experiments will enable researchers to better understand how judges make their decisions, why their current decisions may be erratic and discriminatory, and how pretrial decisions change with the introduction of new tools. A richer understanding of decision-making can allow policymakers to better identify reforms that can aid judges and improve the pretrial system more broadly.

■ *We thank David Arnold, Peter Hull, Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for helpful comments and Dan Ma and Nada Shalash for research assistance. All opinions and errors are our own.*

References

Agan, Amanda Y., and Sonja B. Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133 (1): 191–235.

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30 (2): 175–87.

American Bar Association. 2007. *ABA Standards for Criminal Justice, Third Edition, Pretrial Release.* American Bar Association: Chicago, IL.

Arnold, David, Will Dobbie, and Peter Hull. 2020. "Measuring Racial Discrimination in Bail Decisions." NBER Working Paper 26999.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133 (4): 1885–1932.

Ayres, Ian. 2002. "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy* 4 (1–2): 131–42.

Ayres, Ian. 2010. "Testing for Discrimination and the Problem of Included Variable Bias." Unpublished.

Ayres, Ian, and Joel Waldfogel. 1994. "A Market Test for Race Discrimination in Bail Setting." *Stanford Law Review* 46 (5): 987–1048.

Baradaran, Shima, and Frank McIntyre. 2013. "Race, Prediction, and Pretrial Detention." *Journal of Empirical Legal Studies* 10 (4): 741–70.

Barrett, Joe. 2021. "Some Police Push Back on Bail Reform, Citing Wave of Killings." *The Wall Street Journal,* July 16. https://www.wsj.com/articles/some-police-push-back-on-bail-reform-citing-wave-of-killings-11626441851.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago, IL: University of Chicago Press.

Becker, Gary S. 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101 (3): 385–409.

Board of Governors of the Federal Reserve System. 2018. "Report on the Economic Well-Being of U.S. Households in 2017." Board of Governors of the Federal Reserve System, Washington, DC.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal of Economics* 131 (4): 1753–94.

Canay, Ivan A., Magne Mogstad, and Jack Mountjoy. 2020. "On the Use of Outcome Tests for Detecting Bias in Decision Making." NBER Working Paper 27802.

Cohen, Thomas H., and Brian Reaves. 2007. "Pretrial Release of Felony Defendants in State Courts, State Court Processing Statistics, 1990–2004." Washington, DC: Bureau of Justice Statistics.

Demuth, Stephen. 2003. "Racial and Ethnic Differences in Pretrial Release Decisions and Outcomes: A Comparison of Hispanic, Black, and White Felony Arrests." *Criminology* 41 (3): 873–908.

Demuth, Stephen, and Darrell Steffensmeier. 2004. "The Impact of Gender and Race–Ethnicity in the Pretrial Release Process." *Social Problems* 51 (2): 222–42.

Didwania, Stephanie H. 2020. "The Immediate Consequences of Federal Pretrial Detention." *American Law and Economics Review* 22 (1): 24–74.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Dobbie, Will, and Crystal S. Yang. 2019. *Proposals for Improving the U.S. Pretrial System.* Washington, DC: The Hamilton Project.

Doleac, Jennifer L., and Megan T. Stevenson. 2019. "Algorithmic Risk Assessment in the Hands of Humans." IZA Discussion Paper 12853.

Fishbane, Alissa, Aurelie Ouss, and Anuj K. Shah. 2020. "Behavioral Nudges Reduce Failure to Appear for Court." *Science* 370 (6517).

Frandsen, Brigham R., Lars J. Lefgren, and Emily C. Leslie. 2019. "Judging Judge Fixed Effects." NBER Working Paper 25528.

Gelbach, Jonah. 2021. "Testing Economic Models of Discrimination in Criminal Justice." Unpublished.

Gonnerman, Jennifer. 2015. "Kalief Browder, 1993–2015." *New Yorker,* June 7. http://www.newyorker.com/news/news-desk/kalief-browder-1993-2015.

Grau, Nicolás, and Damián Vergara. 2020. "A Simple Test for Prejudice in Decision Processes: The Prediction-Based Outcome Test." University of Chile Department of Economics Working Paper 493.

Greiner, James, Ryan Halen, Matthew Stubenberg, and Christopher L. Griffin. 2020. "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI." http://a2jlab.org/wp-content/uploads/2020/09/Dane-County-PSA-DMF-RCT-Interim-Report.pdf.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." *Journal of Legal Studies* 45 (2): 471–505.

Hull, Peter. 2021. "What Marginal Outcome Tests Can Tell Us about Racially Biased Decision-Making." NBER Working Paper 28503.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics* 133 (1): 237–93.

Liptak, Adam. 2008. "Illegal Globally, Bail for Profit Remains in U.S." *The New York Times,* January 29. https://www.nytimes.com/2008/01/29/us/29bail.html.

Leslie, Emily, and Nolan G. Pope. 2017. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *Journal of Law and Economics* 60 (3): 529–57.

McKinley, James C., Jr. 2017. "Freed From Custody, Then Accused of Murder: Is Bail System at Fault?" *New York Times*, March 21. https://www.nytimes.com/2017/03/21/nyregion/jose-gonzalez-yadira-arroyo-killing.html.

New York City Criminal Justice Agency, Inc. 2014. *Annual Report 2013.* New York: New York Criminal Justice Agency, Inc.

Ouss, Aurélie, and Megan Stevenson. 2021. "Bail, Jail, and Pretrial Misconduct: The Influence of Prosecutors." http://aouss.github.io/NCB.pdf.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937–75.

Pilnik, Lisa. 2017. *A Framework for Pretrial Justice: Essential Elements of an Effective Pretrial System and Agency.* Washington, DC: The National Institute of Corrections.

Pinto, Nick. 2015. "The Bail Trap." *New York Times Magazine*, August 13. https://www.nytimes.com/2015/08/16/magazine/the-bail-trap.html.

Pretrial Justice Institute. 2018. *Support Grows for Pretrial Justice Reform.* Rockville, MD: Pretrial Justice Institute.

Reaves, Brian A. 2013. "Felony Defendants in Large Urban Counties, 1990–2009." *Bureau of Justice Statistics.*

Schlesinger, Traci. 2005. "Racial and Ethnic Disparity in Pretrial Criminal Processing." *Justice Quarterly* 22 (2): 170–92.

Small, Mario L., and Devah Pager. 2020. "Sociological Perspectives on Racial Discrimination." *Journal of Economic Perspectives* 34 (2): 49–67.

Stevenson, Megan T. 2018. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *Journal of Law, Economics, and Organization* 34 (4): 511–42.

US Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. State Court Processing Statistics. 1990–2009. Felony Defendants in Large Urban Counties. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-06-24. https://doi.org/10.3886/ICPSR02038.v5.

US Department of Justice. 2016. "Brief for the United States as *Amicus Curiae* in Walker v. City of Calhoun, GA, No. 16-10521-HH (11th Cir. Court of Appeals)." https://www.justice.gov/crt/file/887436/download.

US House. 1984. *Bail Reform Act of 1984.* 98th Cong., H.R.5865, Congressional Record.

Walmsley, Roy. 2016. *World Pre-trial/Remand Imprisonment List.* London: Institute for Criminal.

Yang, Crystal S. 2017. "Toward an Optimal Bail System." *New York University Law Review* 92 (5): 1399–1493.

Yang, Crystal S., and Will Dobbie. 2020. "Equal Protection Under Algorithms: A New Statistical and Legal Framework." *Michigan Law Review* 119 (2): 291–396.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 71–96

Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System

Jens Ludwig and Sendhil Mullainathan

The criminal justice system routinely fails at its central mission: delivering justice. Empirical studies reveal a system that is inconsistent in its judgments, mistaken in its predictions, and disparate in its impacts. The same type of defendant handled by different judges is treated very differently, and the same judge treats cases differently from day to day—what behavioral scientists call "noise" (Kahneman, Sibony, and Sunstein 2021). Decisions are often systematically mistaken in ways that could have been better identified in advance: for example, in pre-trial decisions, judges release many high-risk people while simultaneously jailing low-risk people (Kleinberg et al. 2018a). Moreover, certain groups (often disadvantaged in other ways) disproportionately bear the brunt of these problems and receive worse treatment, to the point that one could credibly argue the system is discriminatory against them.

Algorithms have a long history in criminal justice as a potential solution to these problems. Statistical models date back to the 1920s. Explicit guidelines for judges were used even earlier and are themselves primitive algorithms: that is, they are explicit rules for how a judge should decide based on case and defendant characteristics. *In principle*, carefully formed rules provide a way to reduce inconsistency,

■ *Jens Ludwig is the Edwin A. and Betty L. Bergman Distinguished Service Professor, University of Chicago Harris School of Public Policy, Chicago, Illinois. Sendhil Mullainathan is the Roman Family University Professor of Computation and Behavioral Science, University of Chicago Booth School of Business, Chicago, Illinois. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jludwig@uchicago.edu and sendhil@chicagobooth.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.71.

error, and (if constructed with that aim) racial bias (Milgram et al. 2014). It is no surprise, then, that new tools from machine learning have drawn a great deal of interest in criminal justice (Berk 2018). They offer a superior version of what already appealed to many in a very crude form: algorithms trained on large datasets can extract greater predictive signal, and can also rely on inputs that could not have entered simple statistical models or guidelines, such as speech, text, or video. The result is the growing proliferation of algorithms across a wide range of criminal justice applications, as shown in Table 1.

But the optimism for machine learning in criminal justice did not last long. *In practice*, algorithms often proved less helpful than anticipated.[1] In many cases, they were even actively harmful. Some algorithms proved to be no more accurate than the judges whose prediction errors they were purported to correct. Reports emerged of algorithms that were themselves discriminatory, producing racially disparate outcomes at a high enough rate that the phrase "algorithmic bias" has entered the lexicon.[2] The algorithms also introduced new problems of their own, such as a lack of transparency—defendants unable to access the "black boxes" that dictated their fates—and concerns that the system is being depersonalized in a way that compromises due process. The best that could be said, it sometimes seemed, is that at least algorithms are consistent—if inscrutable—in their mistakes.

Why were hopes dashed? One common critique points to features of machine learning itself. According to this argument, the data used to train algorithms are too noisy and biased. The complexity of criminal justice objectives cannot be quantified. These decisions are too important to cede control to black boxes. Consequently, the introduction of algorithms into criminal justice is increasingly viewed as an inherently flawed enterprise. We argue that each of these problems follow from a deeper one. Algorithms fail because of shoddy construction: human decisions about how to build and deploy them is the root cause of problems. Machine learning algorithms in criminal justice are not doomed to fail, but algorithms are fragile: if crucial design choices are made poorly, the end result can be (and is often) disastrous.

One reason for the fragility of algorithms comes from important econometric problems that are often overlooked in building them. Decades of empirical work by economists shows that in almost every data application, the data is incomplete, not fully representing either the objectives or the information that decision-makers possess. For example, judges rely on much more information than is available to algorithms, and judges' goals are often not well-represented by the outcomes provided to algorithms. These problems, familiar to economists, riddle every case where

[1] Throughout the paper we use the term "algorithms" not only in the general sense but also to refer to the more specific end-product of work in the artificial intelligence field of machine learning. We use "artificial intelligence" and "machine learning" interchangeably in what follows and make clear from the context which definition of the term "algorithm" we mean.

[2] "Bias" or "violations of fairness" in social science and legal scholarship usually refers to some combination of disparate treatment, disparate impact, or the principle of fair representation (some also call this "statistical parity"). Computer science, as we discuss below, adds a number of additional definitions. We use the term broadly for most of the paper, but where relevant note which specific definition we mean.

Table 1
Illustrative Applications of Artificial Intelligence in Criminal Justice

Type of application	*Examples*
Investigative/forensic uses	Facial recognition to match closed-circuit television images to mugshots
	Social media image searches to find defendant alibis
	Uses of images for investigations (for example: using backgrounds in image to link suspect to image in child abuse case)
	License plate readers
	Auditing police body-worn camera footage
Detection/monitoring/surveillance	Facial recognition to find lost children, other missing persons
	Gunshot detection
	Chatbots to combat grooming and "sex tourism"
	Closed-circuit television to help airport security decide whom to investigate further
Decision aids	Risk tools for pre-trial release
	Risk tools for diversion decisions
	Risk tools for sentencing
	Risk tools for parole decisions
	Predictive policing (places and times)
	Predictive policing (people)

algorithms are being applied. Another reason is that in criminal justice settings, the algorithm is not the final "decider"—a human is. Building good algorithms requires understanding how human decisions respond to algorithmic predictions. Algorithm builders too often fail to address these types of technical challenges because they haven't had to. Existing regulations provide weak incentives for those building or buying algorithms, and little ability to police these choices.

Economists and other social scientists have a key role to play in building and studying algorithms, because such efforts require econometric, regulatory, and behavioral expertise. The return to such efforts is high: if designed well, algorithms have a chance to undo human fallibility. Algorithms have another benefit—when regulated well, their problems are easier to diagnose and more straightforward to fix than are the problems of human psychology (Kleinberg et. al. 2018c). It is easier to improve fragile algorithms than fallible decision-makers.

We illustrate these ideas for the case of algorithmic bias: why racial disparities arise in algorithms and what can be done about it. We illustrate how poorly built algorithms can exacerbate bias. At the same time, *well-built* algorithms can *reduce* bias. They can, in fact, be a force for social justice. So there is room for cautious optimism: algorithms can still do some good in criminal justice, but only if great care is taken.

The problems and opportunities we highlight for algorithms in criminal justice apply more broadly. Algorithms are increasingly used in a range of areas of interest to economists including the labor market, education, credit, and health care. The issues we raise have equal importance there, and in several cases, have already started to make an appearance. Anyone interested in the effect of algorithms on society has lessons to learn from the criminal justice experience.

Inconsistency, Error, and Discrimination in Judicial Decision-Making

America's criminal justice system is clearly broken. An overwhelming majority of people see the problems and think they must be fixed.[3] The incarceration rate has exploded in a way that has no historical or international precedent (as discussed by Bruce Western in this issue). Nor is the current system, with all of its social costs, providing public safety: America's murder rate far exceeds that of any other high-income nation. Meanwhile, the burden of both crime and incarceration falls disproportionately on minority communities: for example, 70 percent of Black male high school dropouts spend time in prison by their mid-30s.

In this paper, we will focus on inconsistency, error, and discrimination in the criminal justice system. These problems pervade almost every part of the system, ranging from law enforcement to how cases are adjudicated innocent or guilty (plea-bargaining, trials, and other steps) to how people are supervised out in the community on probation or parole. We will focus here on three types of criminal justice decisions that are representative of the broader challenges and substantively important in their own right: pre-trial detention, sentencing, and parole decisions. Given the vast literature, we focus here on selected examples.

The *pre-trial detention* decision occurs soon after an arrest. The defendant must appear in front of a judge within 24 to 48 hours. The judge typically has several choices: release the defendant under their own recognizance (a promise to return for trial); set release with certain conditions, like wearing a location-monitoring device; requiring cash collateral (bail) for release to ensure return to court; or refusal to release the defendant before trial at all. In general, this decision is supposed to depend on the judge's assessment of the defendant's risk to public safety and/or the likelihood that the defendant will appear in court for trial.

The *sentencing decision* occurs when a defendant has been found guilty. This decision will depend on the crime for which the person was convicted but also on the likelihood of future re-offending, as well on other factors like the defendant's remorse and society's sense of just deserts. Depending on the criminal charge, sentencing options could include a fine, probation (the defendant goes free but must report to a probation officer), or detention time either in jail (more common for a misdemeanor charge) or prison (more common for a felony charge).

[3] For an example of such polling data, see Benenson Strategy Group (2017), a survey done for the American Civil Liberties Union.

The *parole decision* arises because historically most defendants sentenced to prison would receive an indeterminate sentence; for example, it might be from four to seven years. After the inmate had served the minimum term, a parole board would then decide when exactly an inmate would go free. Inmates out on parole would typically be required to report periodically to a parole officer. Criteria for parole decisions are similar to those for sentencing but can make use of information about the defendant's behavior in prison as well. The role of parole boards declined in the 1970s with the shift towards determinate sentencing (Kuziemko 2013).

We refer to these three decisions as "judicial decisions" for convenience, recognizing that in practice other criminal justice actors also play a role. Prosecutors, for instance, make recommendations to judges as part of both pre-trial release and sentencing decisions. Prosecutors play a particularly important role for sentencing, given that 90–95 percent of all convictions result from a plea bargain (Devers 2011). The quality of legal representation that defendants receive can vary enormously. For the cases that do make it to trial, a jury may play a role in sentencing. Also, as noted above, while judges often set sentences over a certain range, parole decisions are usually made by a parole board staffed by people who are typically not judges.

The literature has identified three problematic aspects of how decisions are made: for a selective review of some prominent studies, see Table 2. One long-standing concern with these judicial decisions is misprediction: not simply that there are inevitable errors, but that predictions made by judges are systematically mistaken. For example, in the case of sentencing, Gottfredson (1999) asked judges in Essex County, New Jersey in 1977–78 to record their subjective predictions about the recidivism risk of 960 defendants. The correlation between judge predictions and actual recidivism outcomes 20 years later is very modest, on the order of 0.2. These low levels of predictive accuracy also jibe with data from pre-trial release (for example, Jung et al. 2017; Kleinberg et al. 2018a). Concerns with the accuracy of recidivism predictions for parole, which historically have often been made by psychiatrists, dates back at least to the 1940s (Jenkins et al. 1942; Schuessler 1954). More recently, Kuziemko (2013) finds some positive correlation between predictions made by parole boards and recidivism, but Berk (2017) shows there is also substantial misprediction.

Second, judicial decisions are inconsistent in several ways. One way is that they are inconsistent with each other. For example, some judges tend to be "tough" and others "lenient." This discretion was long justified on the basis that judges could then account for the circumstances of each case (Alschuler 1991). But the data shows that even with randomly assigned caseloads, the average level of leniency varies dramatically. Kling (2006), among others, documents this for sentencing, while for pre-trial release decisions, the difference in pre-trial release rates between the most- and least-lenient quintile of judges in New York City was nearly 25 percentage points (Kleinberg et al. 2018a). As one judge complained, "[I]t is obviously repugnant to one's sense of justice that the judgment meted out to an offender should depend in large part on a purely fortuitous circumstance; namely, the personality of the particular judge before whom the case happens to come for disposition" (Diamond and

Table 2
Selected Studies of Inconsistency, Error, and Discrimination in the Criminal Justice System

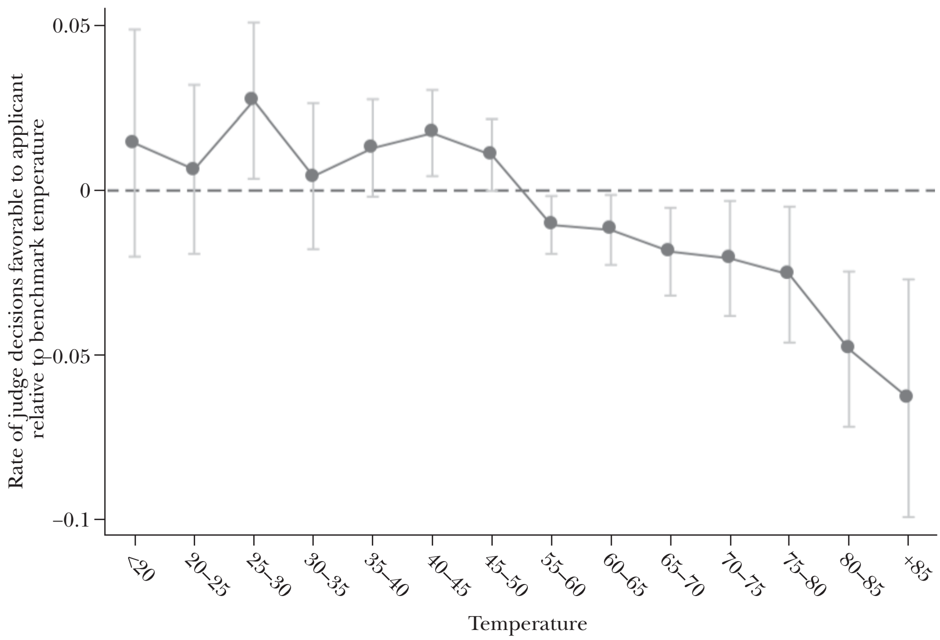
	Error (misprediction)	*Inconsistency*	*Discrimination*
Pre-trial decisions	Judge predictions (implicit in their release decisions) disagree with algorithmic predictions in ways that suggest misprediction (Jung et al. 2017; Kleinberg 2018a)	Judges seeing similar caseloads differ widely in pre-trial decisions, and also deviate from their central tendencies in ways that reduce decision quality (Kleinberg et al. 2018a)	Judges discriminate against minority defendants in an "outcome test" analysis (Arnold, Dobbie, and Yang 2018; Arnold, Dobbie and Hull 2020)
Sentencing decisions	Correlation of judge recidivism predictions with observed recidivism outcomes 20 years later is modest (Gottfredson 1999)	Judges seeing similar case-loads differ widely in prison sentencing (Kling 2006), and are also influenced by irrelevant factors like recent sports-team losses (Eren and Mocan 2018), heat (Heyes and Saberian 2019), or the features of recent cases (Chen, Moskowitz, and Shue 2016)	Judges assign longer sentences to observationally similar minority defendants compared to Whites (Kennedy 2001; Loury 2008; Blumstein 2015)
Parole decisions	Psychiatrist assessment of parolee risk adds little signal beyond structured data (Jenkins et al. 1942); parole predictions disagree with algorithmic predictions in ways that suggest misprediction (Berk 2017)	Parole members pay attention to very different sources of information in their decisions (Ruhland 2020)	Parole may be one of the few parts of the system without substantial racial bias (Anwar and Fang 2015; Mechouan and Sahuguet 2015)

Zeisel 1975, p. 111). Sentencing guidelines were introduced in the 1970s partly to address this problem. But they may have simply shifted discretion to the decisions made by prosecutors about what specific crimes will be charged and what plea-bargain deals will be made (Davis 2005), and of course sentencing guidelines have no effect on pre-trial decisions. For parole decisions, Ruhland (2020) shows parole board members pay attention to very different types of information about a case.[4]

Judges not only differ from each other; they also differ from themselves: the same judge can decide differently on the same case from day to day. For example, Eren and Mocan (2018) show how irrelevant circumstances can skew decisions: Upset losses by the Louisiana State University football team increase the sentences Louisiana judges hand out by about 6 percent—and the effect is larger for judges who are LSU alumni. Heyes and Saberian (2019) show that a 10-degree increase in outdoor temperatures reduces the likelihood an immigration judge rules in favor of

[4] As Kahneman, Sibony, and Sunstein (2021) point out, a more subtle version of across-person inconsistency is when some judges are relatively more lenient on cases of type A and more harsh on cases of type B, while other judges have the reverse pattern.

Figure 1
Variation in Immigration Judge Decisions Favorable to Defendant by Outdoor Temperature



Source: Heyes and Saberian (2019).

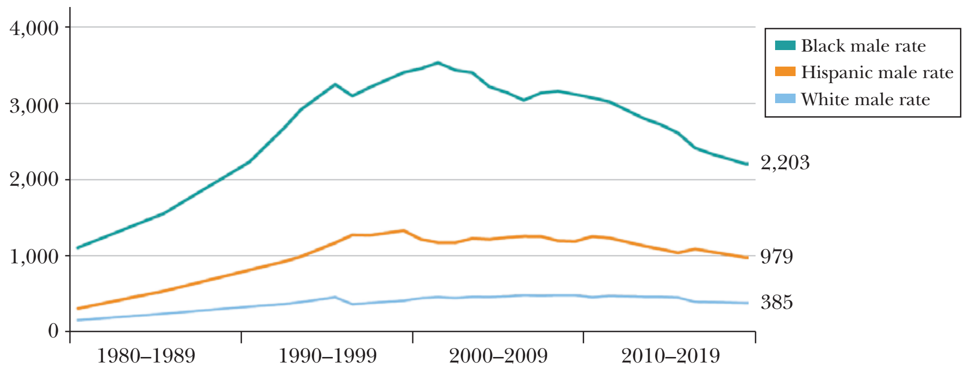
an applicant by 7 percent, as shown in Figure 1. Chen, Moskowitz, and Shue (2016) show that judge decisions in a case depend on the features of other cases the judge just heard. Kleinberg et al. (2018a) present evidence for inconsistency in judicial decisions around pre-trial release. For a defendant, what will happen to you depends on the happenstance of which judge you see and when you happen to see them.

Finally, there are striking racial disparities. For example, African Americans make up 13 percent of the US population, but 26 percent of those who get arrested and 33 percent of those in state prisons.[5] Although disparities in imprisonment have been declining in recent years, they remain substantial (as shown in Figure 2). While disentangling exactly how much of the overall disparity is due to discrimination by the criminal justice system itself is a challenging task, there is little question that some of it is.

As one example of this evidence, Arnold, Dobbie, and Yang (2018) capitalize on the fact that cases are as good as randomly assigned to judges and that judges have systematically different propensities to release defendants pre-trial. They conduct an

[5] Data on the general population is here: https://www.census.gov/quickfacts/fact/table/US/PST045219. Data on state and federal prison inmates is here: https://www.bjs.gov/content/pub/pdf/p19.pdf.

Figure 2
Trends in Incarceration Rates per 100,000 People, by Race and Ethnicity, United States



Source: Bureau of Justice Statistics data analyzed by Statista. https://www.statista.com/chart/18376/us-incarceration-rates-by-sex-and-race-ethnic-origin.

"outcomes test" for marginal defendants. If judges were unbiased, we would expect to see similar re-arrest rates for White and Black defendants with similar probabilities of release. Yet re-arrest rates are lower for Black than White defendants, consistent with judges holding Black defendants to a higher standard. Arnold, Dobbie, and Hull (2020) suggest that around two-thirds of the Black-White disparity in release rates appears to be due to racial discrimination, with statistical discrimination also playing a role. Parole, in contrast, may be one of the few parts of the system where we do not consistently see evidence of substantial racial discrimination (Anwar and Fang 2015; Mechoulan and Sahuguet 2015). For reviews of the larger literature on discrimination in sentencing and many other parts of the justice system, useful starting points include Kennedy (2001), Loury (2008), and Blumstein (2015).

In short, a considerable body of evidence suggests that the criminal justice system is often inconsistent, error-prone, and discriminatory.

The Promise of Artificial Intelligence for Criminal Justice

The limits of human cognition have motivated interest in statistical methods of prediction; in the criminal justice system, "supervised learning" algorithms have become the dominant form of artificial intelligence used. Though the details of building these algorithms can be arcane, they are in essence quite simple. The problem they solve is simple and familiar: given x, predict y (called the "label"). The goal is to look at previous data and form a rule that can be deployed to new situations where x is known, but y is not. Forming those predictions, though, requires large datasets of so-called "labelled observations," where both x and y are available. It is worth noting that every machine learning algorithm is actually two algorithms. The "prediction algorithm" takes as input x and predicts y. It is produced by the "training algorithm," which takes as inputs an entire dataset of (x, y) pairs. In addition, the training algorithms needs an exact objective function specified: more specifically, what is the loss in predicting y incorrectly?

Stated this way, it is clear that familiar economic tools can be viewed as forms of "machine learning." For example, linear regression is one way to predict y from x. The least squares fitting algorithm is, in this case, the "training algorithm" and the "predictor algorithm" is the code which takes the inputs x and multiplies each input by the estimated coefficient. One thing that is new about the current machine learning tools is that they can work with far more complex functional forms and inputs: methodologies like random forests, gradient boosted trees, and neural networks are all examples of non-parametric functional forms which the training algorithm "learns" from the data. Importantly, these tools can also take as a result very novel forms of input: x can be images, audio files, or even video. In this journal, Mullainathan and Spiess (2017) provide an introduction to how machine learning fits in the econometric toolbox.

Importantly, machine learning algorithms fit these complex functions without pre-specification of a functional form by the analyst and while avoiding "over-fitting." A function that fits a specific given dataset as well as possible will inevitably learn more than the general relationship between x and y: it will also be based on statistical noise that is idiosyncratic to that dataset (the overfitting problem), which will, in turn, lead the prediction function to perform poorly on new data. To avoid this problem, these algorithms use sample-splitting techniques where one partition of the data is used for training and model-selection and another for evaluation, ensuring that whatever function is found works well *out*-of-sample.

A well-developed framework in computer science has emerged for building and applying supervised learning algorithms. This framework has enabled breakthroughs in areas like web search, manufacturing, robotics, customer service, automobile safety, and translation. The potential of statistical prediction has only increased over time with the growing availability of "big data" and development of new tools from the artificial intelligence field of machine learning. For excellent reviews at different levels of technical detail, see Athey and Imbens (2019), Berk (2008, 2018), Hastie, Tibshirani, and Friedman (2009), and Jordan and Mitchell (2015), as well as Varian (2014) in this journal.

Building these algorithms requires making key decisions: what outcome to predict, what candidate predictors to make available to the algorithm, and what objective function to provide. For pre-trial release, the relevant outcome is usually guided by state law, usually public safety risk (measured by re-arrest, or measured by re-arrest for violence specifically) and/or flight risk (skipping a required future court case). Typical algorithms used for sentencing and parole instead focus more narrowly on some sort of re-arrest or recidivism risk. Most of these algorithms then use as candidate predictors some combination of the criminal charge for which the person is currently in the justice system, prior criminal record, and a narrow set of demographic factors (usually age, which is legally allowed and predictive of risk given the strong age patterning of criminal behavior, and sometimes gender). Some algorithms can also include factors like employment or some proxy for "community ties" like duration of residence in the area.

These tools differ in important ways in terms of the construction of the functional form that relates the candidate predictors to the outcome of interest. For example, the COMPAS tool that is used for predicting risk of recidivism—and which was the focus of a widely read *Pro Publica* article (Angwin et al. 2016)—is billed as an "evidence-based software product" (http://www.northpointeinc.com/files/downloads/Risk-Needs-Assessment.pdf). But COMPAS is not actually a machine learning tool at all; it seems to be driven instead, as Rudin, Wang, and Coker (2020) note, in large part by human judgments, "a product of years of painstaking theoretical and empirical sociological study" (p. 5). The widely used Public Safety Assessment (PSA) developed by Arnold Ventures for pre-trial release decisions uses a logistic regression to determine the coefficients (weights) that each predictor should get. The tool that the current paper's authors helped develop for use in New York City estimated the relationship between the predictors and the outcome with machine learning, but presents the predictor algorithm to the user as a linear weighted average of predictors to help with interpretability (see also Rudin et al. 2021).

The final ingredient of any algorithm deployed in the criminal justice system is how the results are presented to end-users. Most algorithms map the predictions from the algorithm into recommendations for the final (human) decider. This mapping, typically known as a "decision-making framework," requires making some normative policy judgments about where the right risk thresholds should be to recommend one outcome versus another (like the choice of release versus detain in the pre-trial setting). In practice, those judgments are sometimes made by the algorithm builders alone, sometimes by government agencies, and sometimes through a collaboration. Another question is whether to give the end-user just the recommendations or also the underlying risk predictions, which in principle could help humans learn the algorithm's "confidence" in the recommendation of its decision-making framework (for example, whether a defendant's risk is far from or close to a decision threshold).

These supervised learning algorithms have the potential to improve on human prediction by, for starters, being more accurate. Decades of psychology research show that statistical models, on average, predict more accurately than human beings can in a range of applications (Meehl 1954; Dawes et al. 1989; Grove et al. 2000; Salzinger 2005). That advantage might be even greater today in criminal justice given new supervised learning methods, which allow for increasingly accurate prediction together with the growing availability of larger and larger datasets, which allow for the construction of increasingly accurate algorithms.

Because the predictor algorithm is mechanical, it is necessarily consistent (in the plain English sense of the term). Inputting a given set of predictor-variable values into a predictor algorithm always outputs the same predicted risk. If the human judges and other relevant actors pay attention to the algorithm, there is the potential for an overall reduction in inconsistency (that is, variability in decisions for similar cases) within the justice system.

Finally, statistical models, unlike humans, themselves do not have intrinsic "in-group" preferences—although they can readily acquire such patterns in the training process. What the statistical models learn is a consequence of the training process. As we discuss below, depending on how they are built, their predictions can either mirror historical patterns of discrimination or can undo them.

The prevalence of algorithms within the justice system is hard to determine precisely. This is because data collection and reporting about *anything* in America's justice system is mostly voluntary—even about basic crime statistics—leading to a very underdeveloped criminal justice data infrastructure (Bach and Travis 2021). For pre-trial release, some specific algorithm providers like Arnold Ventures voluntarily share information about use of their tools (at https://advancingpretrial.org/psa/psa-sites). The Arnold tool is used statewide in Arizona, Kentucky, New Jersey, and Utah, in cities like Chicago, Cleveland, Houston, New Orleans, and Pittsburgh, and in a number of suburban and rural counties as well, jurisdictions that together are home to over 66 million people. For sentencing, Stevenson and Doleac (2021) report that algorithms are used in sentencing decisions in a politically, geographically, and demographically diverse set of 28 states; another seven states have at least one county using a risk tool for sentencing. Most states seem to have adopted decision guidelines for parole decisions, if not formal machine learning algorithms, by the mid-1980s (Glaser 1985).

The Disappointing Record of Artificial Intelligence in US Criminal Justice

Against a clear track record of human fallibility and error in the existing criminal justice system, algorithms may seem to offer some hope of improvement. Things have not turned out that way. Supervised learning algorithms and other statistical models in the US criminal justice system have often not only failed to redress problems, they've often created new ones.

The literature analyzing algorithms has focused heavily on documenting racial bias. For example, the widely read *Pro Publica* analysis of the COMPAS risk tool found the tool has a higher false-positive rate in predicting recidivism for Black than White defendants (Angwin et al. 2016). While subsequent research noted the limitations of that specific measure of algorithmic bias,[6] we see examples of algorithms

[6] It is not possible to have both calibration and similar false positive rates with any prediction method (human or algorithmic) in a situation where two groups have different "base rates" for the underlying outcome, unless the prediction method predicts perfectly (Kleinberg, Mullainathan, and Raghavan 2016; Chouldechova 2017).

violating other common definitions of algorithmic fairness as well. For example, *calibration* refers to whether the actual outcomes people experience differ for majority versus protected group members, conditional on the algorithm's risk predictions. This test is complicated by the fact that, for example, we don't observe outcomes for pre-trial defendants who get detained (a point we return to below). With that caveat in mind, we see gender bias in the COMPAS tool (Hamilton 2019). We also see miscalibration by race in the Arnold Ventures Public Safety Assessment, in states like Kentucky, in ways that sometimes create advantages for White defendants (higher crime rate for White than Black defendants at a given risk prediction) and sometimes for Black defendants, as shown in Figure 3. These findings are consistent with evidence of bias in other parts of the justice system that shape the data used by the algorithm, such as police decisions (Fryer 2020; Goncalves and Mello 2021; Hoekstra and Sloan 2020) and jury decisions (Anwar, Bayer, and Hjalmarsson 2012), and consistent with evidence for algorithmic bias in other domains like health (Obermeyer et al. 2019).

Moreover, many of the algorithms that are deployed are either no more accurate than humans or simply have no effect on actual criminal justice outcomes. One review of 19 risk tools used in correctional facilities found them "moderate at best in terms of predictive validity" (Desmarais and Singh 2013; see also Berk 2019). We also see examples where within a few years of adopting algorithms, decisions revert back to the same patterns as before (Stevenson 2018) or fail to meet the objectives policymakers had initially laid out (Stevenson and Doleac 2019) like reduced pre-trial detention.

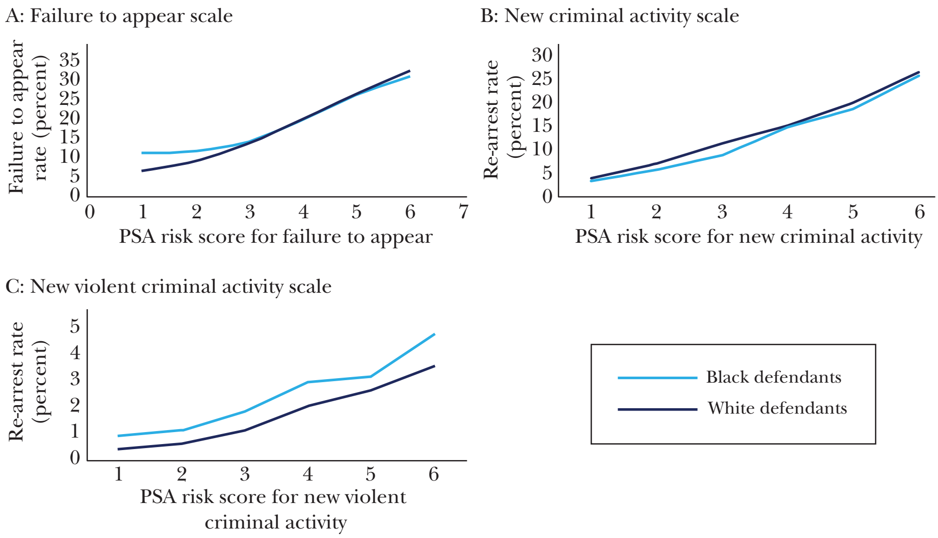
Finally, the adoption of algorithms has also introduced new problems into the criminal justice system, such as limited transparency and concerns about due process. A core value of the American constitutional system, enshrined in the Sixth Amendment to the US Constitution, is the defendant's right to face and confront one's accuser to probe and debate the veracity of the accusations. But many algorithms are not made public, so the defense is deprived of this ability. The Sixth Amendment's "confrontation clause," which was designed reasonably well for the eighteenth century, is severely stretched in the twenty-first. The inability to understand what is happening and why also raises natural concerns about whether the system is treating people in a depersonalized way that compromises Fifth and Fourteenth Amendment due process protections.

Why is Artificial Intelligence Problematic in Practice?

Why have risk tools in the criminal justice system been so disappointing in practice relative to the hoped-for initial promise? The problem is frequently viewed as intrinsic to the machine learning enterprise. Surveys regularly show that the public has a dim view of not just current algorithms, but about their potential to *ever* be useful. For example, one Pew survey found that 58 percent of American adults believe algorithms will inevitably be biased (Smith 2018). Of course many people

Figure 3

Evidence of Mis-calibration in the Public Safety Assessment in Kentucky



Source: DeMichele et al. (2018).

recognize that the alternative to the algorithm—human judgment—can also be biased. So it is revealing that 56 percent of people said in the same survey that they find it "unacceptable" to use algorithms for criminal justice applications like parole. (Majorities also oppose use of algorithms for applications like hiring or credit scoring.) This view is common among experts, too. For example, as one researcher put it: "There's no way to square the circle there, taking the bias out of the system by using data generated by a system shot through with racial bias" (as quoted in Schwartzapfel 2019). Harcourt (2015, p. 237) argued risk tools will "unquestionably aggravate the already intolerable racial imbalance in our prison populations." Similar concerns show up for other problems like accuracy. For example, the belief that reality is easily approximated by a simple combination of one or two factors leads Dressel and Farid (2018) to "cast significant doubt on the entire effort of algorithmic recidivism prediction."

While there are legitimate arguments here, we argue that the overarching reason algorithms perform poorly in practice in the criminal justice system lies elsewhere: many algorithms have been poorly built. Algorithm design, as noted, requires a set of choices and the outcome is highly sensitive to these choices. This creates a fragile process: mistakes in design can lead to consequential errors of the kind we have seen.

The mistakes are, perhaps first and foremost, basic technical ones that arise in working with messy data generated by past human decisions. If econometrics = statistics + human agency, most algorithms are built not through an econometric approach but through a statistical one that ignores key aspects of this messiness. The development of these tools also too often ignores a key sociological challenge: Algorithms don't make decisions, people do. Regulatory failures provide the underlying reason for the persistence of both of these types of technical failures: No safeguards are in place currently to stop inadequately built algorithms from being deployed.

Badly Built Algorithms

Many algorithms cause harm because they have not been constructed to solve two types of econometric challenges. The first is the potential for misalignment between algorithmic objectives and human decision-maker objectives, a problem that is rampant in criminal justice and also shows up in many other areas as well such as child welfare screening (Coston, Chouldechova, and Kennedy 2020). The second is that the data we have are filtered by past decisions of humans who may see things about cases that are not captured in the data.

Nearly all machine learning algorithms simply predict outcomes. But a judicial decision often depends on more—sometimes much more—than the prediction of a single outcome. By assuming that all that matters is the outcome being predicted, an algorithm can wind up leaving out many of the elements the decision-maker cares about. We call that problem *omitted payoff bias* (Kleinberg et al. 2018a).

To see the problem, note that artificial intelligence tools are regularly built for all three judicial decisions we study here: pre-trial, sentencing, and parole. An implicit assumption is that prediction of an outcome like re-arrest or recidivism is equally useful in each case, but in fact, the role that prediction plays in the decision is quite different. For sentencing, for example, countless examples make clear that the objective function of real-world judges is richer than this; it can also include defendant circumstances, personal culpability, remorse, and society's sense of just deserts. Thus, decision-makers are receiving predictions only for a subset of what matters for their decision, creating risk of distorting the decision outcome. In contrast, pre-trial release decisions are supposed to depend on a narrower set of criteria: the judge's prediction of the defendant's flight or public safety risk. A recidivism predictor is better suited for pre-trial decisions than for sentencing because what the algorithm is specifically predicting is better aligned with the judge's objectives. This difference helps to explain why so much recent work on algorithms in the criminal justice system has been focused on the pre-trial release decision.[7] For an economist, an obvious way to address this problem is to inform the algorithm design with a model of the human decision-maker's actual objective function.

A related danger lies in mistakenly concluding the algorithm improves upon human-only decisions because it is better on one dimension, even if it ignores other dimensions the human decision-makers may care about. For example, in the case of pre-trial release tools, Kleinberg et al. (2018a) build an algorithm to predict failure

[7] For example, see Anderson et al. (2019); Angelino et al. (2017); Berk, Sorenson, and Barnes (2016); Corbett-Davies et al. (2017); Cowgill and Tucker (2017); Jung et al. (2017); Kleinberg et al. (2018a); Stevenson (2018); Jung, Goel, and Skeem (2020); and Wang et al. (2020).

to appear in court, the outcome which local law in New York state says should be the focus of risk considerations in pre-trial decisions. But they then show that the algorithm also dominates judge decisions on *other* outcomes that could potentially enter the judge's objective function in practice like re-arrest risk, risk of any violence, or serious violence specifically or, as discussed further below, racial disparities. However, this sort of comprehensive assessment of multiple objectives is not yet standard practice in the field.

The second econometric complication that arises in constructing algorithms also stems from the basic fact that there is a human in the loop in criminal justice applications, namely: we are selectively missing some of the data we need to evaluate the algorithms. Conceptually, the problem is to compare status quo decision-making with the decisions that would happen with an algorithm in the decision loop. However, the data available to estimate that counterfactual are generated by past judicial decisions. If the algorithm recommends pre-trial detention of a defendant that past judges had detained pre-trial, constructing counterfactual arrests is easy because the number of crimes committed before trial is, by construction, zero in both cases. But if the algorithm recommends release of a defendant that judges detained, what do we do? We are missing a measure of the defendant's behavior if released in that case.[8]

We cannot simply impute the missing outcomes for these defendants by looking at the outcomes of released defendants who appear similar on measured variables. There may be cases where judges make inconsistent decisions, as we discussed earlier, but in addition, the judge may have access to information not captured by the algorithm. The presence of *unobserved* variables means two observations we *think* are comparable based on our data may not actually be comparable. We call this the *selective labels problem* (Kleinberg et al. 2018a).

Econometrics has tools to address this issue. For example, economists studying pre-trial judicial decisions have used the fact that judges vary in their leniency rates, and that sometimes defendants are as good as randomly assigned to judges (Kleinberg et al. 2018a; Arnold, Dobbie, and Hull 2020, 2021; Rambachan and Roth 2020; Rambachan 2021). For example, we can take the caseload of more lenient judges, use the algorithm to select the marginal defendants to detain to get down to the detention rate of stricter judges, then compare the observed crime under the simulated lenient judges plus algorithm detention rule compared to the observed stricter judges' crime rate. This allows us to evaluate the algorithm's performance, focusing only on the part of the counterfactual estimation problem (contracting or shrinking the released set) where the selective labels problem is not binding. This sort of exercise confirms that algorithms are indeed able to predict risk much more accurately than can human judges.

[8] There is another problem here from missing outcome data for defendants the judge detains, which is that we can only build an algorithm using data from released defendants. That problem could reduce accuracy of the algorithm when applied to the full set of defendants who come in for pre-trial or sentencing hearings in the future. That possibility just further reinforces the importance of being able to solve the evaluation challenge mentioned above to determine whether the built algorithm really could predict more accurately than do judges.

This selective-labels problem shows up in any situation where the human decision affects the availability of the label, such as hiring where we only observe performance on the job for the employees that a firm decided to hire (for example, Hoffman, Kahn, and Li 2018). In criminal justice applications, it shows up repeatedly and sometimes in different forms. For example, in predictive policing the crime data we have for evaluating the potential performance of any new algorithm are generated by past policing decisions about where to deploy resources. In this approach, the outcome values are contaminated by the treatment effects caused by past decisions (Mullainathan and Obermeyer 2017). The evaluation tools provided by econometrics have not yet diffused into standard operating practice for algorithms before they are deployed at scale.

Human Plus Machine Challenges

A second type of technical challenge that real-world algorithms often fail to address adequately stems from the fact that the *algorithm* does not decide. Humans remain in the loop as the ultimate decision-makers. Thus, any successful algorithmically informed system will need to not only design the algorithm correctly, but also understand and allow for how humans use these algorithms in practice.

A common approach is to assume that because algorithmic predictions can be more accurate than those of humans *on average*, the goal should just be to get the human to follow the algorithm's recommendations as often as possible. The assumption that the algorithm is (almost) always right is reflected in the increasingly common term "algorithm aversion"—the behavioral science description for people's reluctance to always follow the recommendation of a prediction tool (Dietvorst, Simmons, and Massey 2015). Similarly, when economists and others have focused on evaluating deployments of artificial intelligence in criminal justice, they often focus on the "problem" of the human not following the algorithm enough.

But simply getting the human to mindlessly follow the algorithm as often as possible is not the right goal, not only because few humans will love the idea of effectively being replaced like this, but also because it need not be the social welfare-maximizing approach. While an algorithm does indeed have an advantage over humans in being able to access a large number of administrative data (a "longer" dataset) to form predictions, humans often have access to data the algorithm does not (a "wider" dataset). This raises the possibility that at least in some cases the human can have an advantage over the algorithm (for example, De-Arteaga, Fogliato, and Chouldechova 2020). Determining when the human should follow the algorithm's prediction, or not, is what we call the *override problem*.

Consider a situation with two sources of information for making a decision about pre-trial release: information observable to both the algorithm and the judge, and information unobservable by the algorithm but observable by the judge. In this setting, consider two possible scenarios that might arise. In the first scenario, the judge using the additional information always estimates more accurately, which in some cases leads to correcting errors that would have been made by the algorithm.

That is, when the algorithm and the judge disagree, the judge is correct to override the algorithm—if the algorithm had the additional information, it would agree with the judge's decision. In the second scenario, the judge uses the additional information in a way that always leads to an incorrect decision: that is, if the algorithm had full information on not just its usually observed data but also the unobserved information usually seen just by the judge, it would still disagree with the judge. In this scenario, when the algorithm and the judge disagree, the algorithm is correct even based on limited information—because the judge draws the wrong inference from fuller information.

Solving the override problem raises new frontier-science challenges that the omitted payoffs and selective labels problems typically do not. The deep problem that has not yet been fully figured out is to understand the contexts in which humans and machines working together might do better than either alone (for example, Salzinger 2005; Jussupow, Benbasat, and Heinzl 2020). Solving the override problem requires not just helping judges use their information as well as possible, but also helping them learn where they have comparative advantage over the algorithm and vice versa. That, in turn, requires figuring out ways of helping judges better understand the algorithm, a focus of computer science work on interpretable algorithms.

It's worth noting that what it even means for something to be interpretable as "an explanation" is unclear. Psychology shows that people find even vacuous explanations acceptable if they simply begin with the word "because." For example, Langer, Blank, and Chanowitz (1978) show that study subjects are more likely to let someone cut in line in front of them at the photocopy machine when the person offers a reason ("because I'm in a hurry") than when they don't. But they're equally likely to let someone cut in line with a real reason as with the vacuous veneer of a reason ("because I need to make copies"). Identifying ways of communicating the process and recommendations of artificial intelligence to humans is as much about understanding the human as it is about the algorithm. More fundamentally, given the importance of due process, solving this problem is essential: when a person is detained or imprisoned based in part on an algorithm's recommendations, "it's a complicated black box" is not an acceptable answer for why.

The fact that algorithms often fail in criminal justice because of the behavior of the human users rather than the artificial intelligence technology itself means that social science will inevitably have an important role to play in solving these problems. Progress on these issues will require creativity in data collection of the sort at which applied economists have become adept, combined with the ability of artificial intelligence methods to make use of unstructured data sources that may help capture the sources of the judge's private signal such as text (courtroom transcripts) or images (perhaps use of video from the courtroom).

Evidence that progress on these human plus machine challenges is possible comes from the progress that fields other than criminal justice have made. For example, to help radiologists detect breast cancer from mammograms Jiang et al. (1999) not only built an algorithm but designed a user interface that presented the doctors with the information in ways they are used to seeing, which in turn, improved diagnostic outcomes. Tschandl et al. (2020) tested multiple user interfaces for the algorithm and came up with an algorithm-human combination that leads to better diagnosis than either the algorithm or the doctor alone (also, see the review in Doi 2007). The fact that this type of progress shows up in medicine, but not in criminal justice, is no accident—as we discuss next.

Inadequate Procurement and Regulation

Why have so many real-world algorithms failed to deal with problems like omitted payoff bias, the selective labels problem, and the override problem? The answer, in short, is that they have not had to. The parties involved in building and deploying algorithms lack either the information or motivation needed to solve those problems, and there are no corrective mechanisms to prevent the flawed algorithms that result from being deployed widely.

Part of the problem is that algorithms used by criminal justice agencies are often not built by those agencies. Vendors can often have asymmetric information with regard to buyers, as well as potentially divergent interests—ideas that are very familiar to economists. With algorithms in the criminal justice system specifically, the vendors often have incorrectly specified the problem to be solved. For example, the allocation of social programs for those in the criminal justice system is often guided by algorithms that predict risk of crime involvement (a standard predictive-inference problem) rather than by the predicted benefit from intervention (a causal-inference problem). Even if the problem is correctly specified, the algorithm's ability to achieve that goal is unclear because few algorithms are properly evaluated prior to deployment. But the buyers don't have the ability to tell, and the result will be a system that does not perform as hoped.

We often rely on regulation to deal with underinformed consumers, but (as is often the case with new technologies) the law and larger regulatory apparatus is still catching up to the ways in which artificial intelligence can cause harm. For example, in health, the Food and Drug Administration requires new medicines or medical devices to be rigorously evaluated through a series of randomized controlled trials before they are deployed. No similar requirement currently exists for algorithms.

The limitations of current algorithmic regulations are not limited to procurement. For example, current discrimination laws are designed to deal with human bias, but fail to deal with how algorithms discriminate (Kleinberg et al. 2018c). Discrimination law for humans focuses on ensuring that people don't pay attention to protected group characteristics. The human brain is the ultimate black box, so we can't tell when a person would use such characteristics to enhance versus detract from accepted societal goals. In contrast, as we discuss further below, for algorithms the use of protected group characteristics can actually help undo bias (Dwork et al. 2012; Kleinberg et al. 2018b; Goel et al. 2021). Discrimination law built for humans is silent on what we outside observers need to monitor algorithms for bias, such as access to data and the predictor algorithm for "fairness audits" and improved transparency in general (Rudin, Wang, Coker 2020).

Algorithmic Bias

Much of the public debate around algorithms explicitly or implicitly assumes that their problems are intrinsic to the underlying technology. Our argument instead is that the problems with algorithms stem not from something intrinsic to artificial intelligence but instead from human decisions about how to construct, evaluate, deploy, and regulate these tools, as shown in Table 3. Indeed, we argue that there are principled ways to address the problems with these underlying human decisions. To illustrate that argument, we consider the case of algorithmic bias.

Under our framework, algorithmic bias is largely an example of omitted payoff bias. Society has a strong social preference for fairness (as well as predictive accuracy). Yet the algorithm builder may ignore this preference and focus only on predictive accuracy. As a result, the wrong data can be used (for example, an algorithm that predicts an outcome like arrests for low-level offenses where officer discretion is high, hence risk of bias is high); tools are evaluated using the wrong outcome criteria (for example, by accuracy alone versus a comparison along multiple dimensions that includes fairness as one); or how the algorithm's output is presented to the judge (for example, if many other factors matter to the judge, providing recommendations rather than the specific narrow predictions can be misleading).

In contrast, once fairness objectives are recognized, they can be incorporated. Concerns about bias in data can lead the algorithm builder to focus on using data on more serious rather than less serious offenses, if discretion (and hence bias) is attenuated with the former, or focusing on convictions over arrests. Different machine learning models can have similar rates of overall predictive accuracy but differ in their predictions for specific cases (the so-called "Rashomon effect"), and so can lead to different implications for fairness objectives (for example, Coston, Rambachan, and Chouldechova 2021).

There are also additional design choices that could be made to improve algorithmic fairness, even if some of them are currently prohibited by laws designed to deal with how humans rather than algorithms discriminate (Kleinberg et al. 2018c; Goel et al. 2021). For example, allowing a properly built algorithm to access information about protected-group membership can help *undo* the effects of bias in the underlying data (Dwork et al. 2012; Kleinberg et al. 2018b). As an example, imagine that in some city, half of all arrests of minority residents are false arrests (the person did not actually commit a crime), while none of the arrests of White residents are. In that case, an algorithm blinded to group membership has no choice but to treat each arrest as equally informative about risk of flight or re-arrest. In contrast, an algorithm that knows a defendant's race or ethnicity has the potential to learn that arrests to minority residents contain less "signal" about future outcomes than do arrests to White residents and could estimate a different arrest-to-risk relationship for each group and so undo some of the bias baked into the underlying arrest data. A similar approach would involve setting different risk thresholds for release for different groups.

Not only is fairness too often ignored, the variability of fairness preferences are also ignored. After all, the most widely used risk tools were built for use in multiple

Table 3
Common Concerns with Algorithms as Explained by Our Framework

Concern	*Example of failure to solve technocratic problem (omitted payoffs, selective labels, override)*	*Example of regulation/procurement problem*
Ineffectiveness	Inaccurate algorithm mistakenly evaluated to be effective because of failure to deal with selective labels	Algorithm not required to be adequately evaluated before deployment
Transparency		Algorithm not made public because buyer and regulations did not require it
Due process/ depersonalization	Judges may override highly accurate algorithms in ways that reduce differentiation across defendants	Algorithms with low predictive accuracy fail to adequately distinguish among defendants in pre-trial release decisions
Fairness	Algorithm built without adequate attention to human decision-maker's equity objectives	Procurement of biased algorithms when unbiased algorithms for same purpose are available

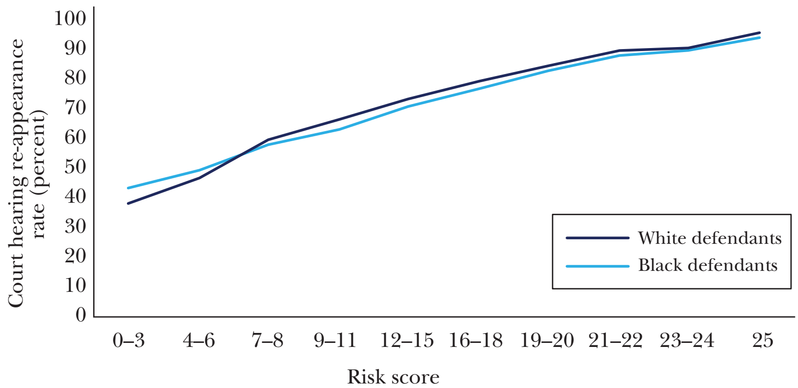
jurisdictions; they were not designed to reflect the specific equity or other preferences of any particular place. Put differently, algorithms (unlike humans) come with "equity knobs"—the ability to make adjustments in response to the specific equity objectives of a given policymaker.

Proof-of-concept of what is possible from accounting for equity preferences comes from an algorithm to inform pre-trial release decisions in New York City that one of our research centers (the University of Chicago Crime Lab) helped construct. New York was one of the first places in the United States to implement a pre-trial risk tool back in the 1960s, as part of the Vera Institute of Justice's Manhattan Bail Project. The new tool that our team worked to develop with Luminosity and New York's Criminal Justice Agency was implemented in November 2019. The previous tool that had been in use since 2003 (!) showed signs of miscalibration by race. In contrast, the new tool that our team built meets the calibration test, as seen in Figure 4.

Perhaps even more important than the algorithm's statistical properties are its effects on decision outcomes, as shown in Table 4. The older tool recommended release for 32 percent of Black defendants and 41 percent of White defendants. New York City government set the new release thresholds based on estimates for how much higher the release rate could go without increasing failure-to-appear rates, where the possibility of increasing release without increasing the risk of failure-to-appear for a future court proceeding comes from better prioritizing the truly high risk for detention. As shown in Table 4, the new tool our team helped build recommends for release 83.9 percent of Black defendants and 83.5 percent for White defendants—a large absolute gain in release rates for both groups, and a reduction in the racial gap from nine percentage points down to effectively zero. That is, our new tool meets not only the calibration definition for algorithmic fairness, but even the more stringent (and more controversial) definition of "statistical

Figure 4

Calibration Test of New York City's New Release Assessment—Reappearance Rates by Predicted Risk Bin, by Race/Ethnicity



Source: Luminosity and University of Chicago Crime Lab (2020).

Table 4

Results from Algorithm for Pre-trial Release Decisions in New York City

	Release recommendations under old tool	*Release recommendations under new tool*	*Judge release decisions under new tool (2019–20 data)*
Black defendants	31.7%	83.9%	69.4%
White defendants	41.1%	83.5%	72.0%
Black-White gap	9.4 percentage points	0.4 percentage points	2.6 percentage points

Source: Peterson (2020). The new algorithmic tool was built by the University of Chicago Crime Lab in partnership with Luminosity and the NYC Criminal Justice Agency.

parity" (Hertweck, Heitz, and Loi 2021). To underscore the point that, at the end of the day, the justice system is more about the humans than the technology, Table 4 also shows what ultimately happened in practice when the tool was deployed: the human judges took release recommendations that were similar across race groups and turned them into a three-point gap in favor of Whites (Peterson 2020).

Our key point is that with the right motivations for the human algorithm builders and deployers, algorithms have the *potential* not only to avoid bias, but even to be a force for social justice. We see other examples in policing, for instance, where incorporating fairness objectives changed algorithmic outcomes for hiring decisions by the Los Angeles Police Department (Ridgeway 2013) and, according to evidence from a randomized trial, led to a predictive policing tool that helped reduce crime without increasing overall arrests or the racial composition of those arrested (Mohler et al. 2015; Brantingham, Valasik, and Mohler 2018). Examples of how to incorporate fairness objectives into algorithms, and examples of how doing so can lead to gains relative to the status quo, show up in many other domains of interest to economists as well, such as hiring (Li, Raymond, and Bergman 2020), lending (Bartlett et al. 2019), housing (Ross and Yinger 2002), and health (Obermeyer et al. 2019).

Racial bias provides a useful contrast between human and algorithmic decision-making. Discrimination by people is hard to discover (Charles and Guryan, 2011). Once found, it is hard to fix. As an example, intricate hiring audits are needed to uncover bias in resume screening, and even despite the widespread dissemination of those findings, little has changed over the last two decades (Bertrand and Mullainathan 2004; Kline, Rose, and Walters 2021). Algorithms can, with the right transparency measures, be more straightforwardly audited and adjusted. With the right motivations and regulations in place, algorithmic bias can be easier to find and fix than human bias (Kleinberg et. al. 2018c).

Conclusion

Very often the discussion of algorithms happens in a vacuum. For many social systems, including but not limited to criminal justice, we cannot understand the algorithms without understanding the human beings. Humans set the benchmark for algorithms through their existing decisions. Humans produce the data that the algorithm uses. Humans build and deploy the algorithm. Viewed this way, we can see that algorithms cannot be expected to be an automatic panacea for all the problems of our criminal justice system. Algorithms can be, and too often in practice are, deeply problematic.

But they need not be. Designed correctly, they offer a potential remedy for human fallibility. The challenge to overcome is that algorithms themselves are fragile, extremely sensitive to design choices. Those choices are made and the resulting algorithms are built, deployed, and procured by a social system riddled with the very problems we seek to address, a system that has been designed and implemented by fallible humans. These problems are complex but not hopeless. Economists and other social scientists have an important role to play in ensuring that algorithms do no harm and even do social good.

■ *Thanks to Amanda Agan, Charles Brown, Alexandra Chouldechova, Philip Cook, Amanda Coston, Dylan Fitzpatrick, Barry Friedman, Jonathan Guryan, Peter Hull, Erik Hurst, Sean Malinowski, Nina Pavcnik, Steve Ross, Cynthia Rudin, Greg Stoddard, Cass Sunstein, Timothy Taylor, Heidi Williams, and Morgan Williams for valuable comments, and to Kristen Bechtel, Megan Cordes, Ellen Dunn, Rowan Gledhill, and Elizabeth Rasich for their assistance. All opinions and any errors are of course our own. The authors gratefully acknowledge support for the construction of the New York City algorithm discussed in the paper from the non-profit Criminal Justice Agency to the University of Chicago Crime Lab, and for support for the writing of this paper from the Sloan Foundation and the Center for Applied Artificial Intelligence at the University of Chicago Booth School of Business.*

References

Alschuler, Albert W. 1991. "The Failure of Sentencing Guidelines: A Plea for Less Aggregation." *University of Chicago Law Review* 58 (3): 901–51.

Anderson, Chloe, Cindy Redcross, Erin Valentine, and Luke Miratrix. 2019. *Evaluation of Pretrial Justice System Reforms that Use the Public Safety Assessment: Effects of New Jersey's Criminal Justice Reform*. New York City: MDRC.

Angelino, Elaine, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin. 2017. "Learning Certifiably Optimal Rule Lists for Categorical Data." Paper presented at KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, August 13–17.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias". *ProPublica*, May 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127 (2): 1017–1055.

Anwar, Shamena, and Hanming Fang. 2015. "Testing for Racial Prejudice in the Parole Board Release Process: Theory and Evidence." *Journal of Legal Studies* 44 (1).

Arnold, David, Will S. Dobbie, and Peter Hull. 2020. "Measuring Racial Discrimination in Bail Decisions." NBER Working Paper 26999.

Arnold, David, Will Dobbie, and Peter Hull. 2021. "Measuring Racial Discrimination in Algorithms." *AEA Papers and Proceedings* 111: 49–54.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133 (4):1885–1932.

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods that Economists Should Know About." *Annual Review of Economics* 11: 685–725.

Bach, Amy, and Jeremy Travis. 2021. "Don't Ignore the Infrastructure of Criminal Justice." *The Hill*, August 16. https://thehill.com/blogs/congress-blog/politics/568063-dont-ignore-the-infrastructure-of-criminal-justice.

Barlett, Robert, Adair Morse, Richard Stanton and Nancy Wallace. 2019. "Consumer-Lending Discrimination in the Fintech Era." NBER Working Paper 25943.

Benenson Strategy Group. 2017. "2017. ACLU National Survey." https://www.aclu.org/report/smart-justice-campaign-polling-americans-attitudes-criminal-justice.

Berk, Richard A. 2008. *Statistical Learning from a Regression Perspective*, Vol. 4. New York: Springer.

Berk, Richard A. 2017. "An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism." *Journal of Experimental Criminology* 13: 193–216.

Berk, Richard A. 2018. *Machine Learning Risk Assessments in Criminal Justice Settings*. New York: Springer.

Berk, Richard A. 2019. "Accuracy and Fairness for Juvenile Justice Risk Assessments." *Journal of Empirical Legal Studies* 16 (1): 175–94.

Berk, Richard A., Susan B. Sorenson, and Geoffrey Barnes. 2016. "Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions." *Journal of Empirical Legal Studies* 13 (1): 94–115.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Blumstein, Alfred. 2015. "Racial Disproportionality in Prison." In *Race and Social Problems*, edited by Ralph Bangs and Larry E. Davis, 187–93. New York: Springer.

Brantingham, P. Jeffrey, Matthew Valasik, and George O. Mohler 2018. "Does Predictive Policing Lead to Biased Arrests? Results from a Randomized Controlled Trial." *Statistics and Public Policy* 5 (1).

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. "Studying Discrimination: Fundamental Challenges and Recent Progress." *Annual Review of Economics* 3: 479–511.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *Quarterly Journal of Economics* 131 (3): 1181–1242.

Chouldechova, Alexandra. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." *Big Data* 5 (2): 153–63.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. "Algorithmic Decision

Making and the Cost of Fairness." Paper presented at the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, August 13–17.

Coston, Amanda, Alexandra Chouldechova, and Edward H. Kennedy. 2020. "Counterfactual Risk Assessments, Evaluation, and Fairness." Paper presented at 2020 Conference on Fairness, Accountability and Transparency, Barcelona, January 27–30.

Coston, Amanda, Ashesh Rambachan, and Alexandra Chouldechova. 2021. "Characterizing Fairness over the Set of Good Models under Selective Labels." arXiv pre-print.

Cowgill, Bo, and Catherine Tucker. 2017. "Algorithmic Bias: A Counterfactual Perspective." https://bitlab.cas.msu.edu/trustworthy-algorithms/whitepapers/Bo%20Cowgill.pdf.

Davis, Angela J. 2005. "The Power and Discretion of the American Prosecutor." *Droit et Cultures* 49: 55–66.

Dawes, Robyn M., David Faust, and Paul E. Meehl. 1989. "Clinical versus Actuarial Judgement." *Science* 243 (4899): 1668–74.

De-Arteaga, Maria, Riccardo Fogliato, and Alexandra Chouldechova. 2020. "A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores." Paper presented at ACM Conference on Human Factors in Computing Systems, Honolulu, April 25–30.

Desmarais, Sarah L., and Jay P. Singh. 2013. "Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States." https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf.

Devers, Lindsey. 2011. *Plea and Change Bargaining*. Washington, DC: Bureau of Justice Assistance, US Department of Justice.

Diamond, Shari Seidman, and Hans Zeisel. 1975. "Sentencing Councils: A Study of Sentence Disparity and Its Reduction." *University of Chicago Law Review* 43 (1): 109–49.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." *Journal of Experimental Psychology: General* 144 (1): 114–26.

DeMichele, Matthew, Peter Baumgartner, Michael Wenger, Kelle Barrick, and Megan Comfort, and Shilpi Misra. 2018. "The Public Safety Assessment: A Re-Validation and Assessment of Predictive Utility and Differential Prediction by Race and Gender in Kentucky." http://dx.doi.org/10.2139/ssrn.3168452.

Doi, Kunio. 2007. "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential." Computerized Medical Imaging and Graphics 31 (4–5): 198–211.

Dressel, Julia, and Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." Science Advances 4 (1).

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold and Richard Zemel. 2012. "Fairness through Awareness." Paper presented at 3rd Innovations in Theoretical Computer Science Conference, Cambridge, MA, January 8–10.

Eren, Ozkan, and Naci Mocan. 2018. "Emotional Judges and Unlucky Juveniles." *American Economic Journal: Applied Economics* 10 (3): 171–205.

Fryer, Roland G., Jr. 2020. "An Empirical Analysis of Racial Differences in Police Use of Force: A Response." *Journal of Political Economy* 128 (10): 4003–4008.

Glaser, Daniel. 1985. "Who Gets Probation and Parole: Case Study versus Actuarial Decision Making." *Crime and Delinquency* 31 (3): 367–78.

Goel, Sharad, Ravi Shroff, Jennifer Skeem, and Christopher Slobogin. 2021. "The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment." In *Research Handbook on Big Data Law*, edited by Roland Vogl. Cheltenham: Edward Elgar Publishing.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–41.

Gottfredson, Don M. 1999. *Choosing Punishments: Crime Control Effects of Sentences.* Washington, DC: US Department of Justice.

Grove, William, David Zald, Boyd Lebow, Beth Snitz, and Chad Nelson. 2000. "Clinical versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12 (1): 19–30.

Hamilton, Melissa. 2019. "The Sexist Algorithm." *Behavioral Sciences & The Law* 37 (2): 145–57.

Harcourt, Bernard E. 2015. "Risk as a Proxy for Race: The Dangers of Risk Assessment." *Federal Sentencing Reporter* 27 (4): 237–43.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer.

Hertweck, Corinna, Christoph Heitz, and Michele Loi. 2021. "On the Moral Justification of Statistical Parity." Paper presented at FAccT '21, Virtual Event, Canada, March 3–10.

Heyes, Anthony, and Soodeh Saberian. 2019. "Temperature and Decisions: Evidence from 207,000 Court Cases." *American Economic Journal: Applied Economics* 11 (2): 238–65.

Hoekstra, Mark, and CarlyWill Sloan. 2020. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." NBER Working Paper 26774.

Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li. 2018. "Discretion in Hiring." *Quarterly Journal of Economics* 133 (2): 765–800.

Jenkins, R.L., Henry Harper Hart, Philip I. Sperling, and Sidney Axelrad. 1942. "Prediction of Parole Success: Inclusion of Psychiatric Criteria." *Journal of Criminal Law and Criminology* 33: 38–46.

Jiang, Yulei, Robert Nishikawa, Robert Schmidt, Charles Metz, Maryellen Giger, and Kunio Doi. 1999. "Improving Breast Cancer Diagnosis with Computer-Aided Diagnosis." *Academic Radiology* 6 (1): 22–33.

Jordan, Michael I., and Tom M. Mitchell. 2015. "Machine learning: Trends, Perspectives, and Prospects." *Science* 349 (6245): 255–60.

Jung, Jongbin, Connor Concannon, Ravi Shroff, Sharad Goel, and Daniel G. Goldstein. 2017. "Simple Rules for Complex Decisions." *arXiv preprint arXiv:1702.04690.*

Lin, Zhiyuan, Jongbin Jung, Sharad Goel, and Jennifer Skeem. 2020. "The Limits of Human Predictions of Recidivism." *Science Advances* 6 (7).

Jussupow, Ekaterina, Izak Benbasat, and Armin Heinzl. 2020. "Why Are We Averse towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion." Paper presented at the 28th European Conference on Information Systems, virtual event, June.

Kahneman, Daniel, Oliver Sibony, and Cass R. Sunstein. 2021. *Noise: A Flaw in Human Judgement.* New York City: Little, Brown, and Company.

Kennedy, Randall. 2001. "Racial Trends in the Administration of Criminal Justice." *America Becoming: Racial Trends and Their Consequences* 2: 1–20.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018a. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics* 133 (1): 237–93.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. 2018b. "Algorithmic Fairness." *AEA Papers and Proceedings* 108: 22–27.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2018c. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis* 10: 113–74.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent Trade-Offs in the Fair Determination of Risk Scores." *arXiv preprint arXiv:1609.05807.*

Kline, Patrick M., Evan K. Rose, and Christopher R. Walters. 2021. "Systemic Discrimination among Large US Employers." NBER Working Paper 29053.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76.

Kuziemko, Ilyana. 2013. "How Should Inmates Be Released from Prison? An Assessment of Parole versus Fixed Sentence Regimes" *Quarterly Journal of Economics* 128 (1): 371–424.

Langer, Ellen J., Arthur Blank, and Benzion Chanowitz. 1978. "The Mindless of Ostensibly Thoughtful Action: The Role of 'Placebic' Information in Interpersonal Interaction." *Journal of Personality and Social Psychology* 36 (6): 635–42.

Li, Danielle, Lindsey R. Raymond, and Peter Bergman. 2020. "Hiring as Exploration." NBER Working Paper 27736.

Loury, Glenn C. 2008. *Race, Incarceration, and American Values.* Cambridge, MA: MIT Press.

Luminosity and University of Chicago Crime Lab. 2020. *Updating the New York City Criminal Justice Agency Release Assessment.* New York: Criminal Justice Agency.

Mechoulan, Stéphane, and Nicolas Sahuguet. 2015. "Assessing Racial Disparities in Parole Release." *Journal of Legal Studies* 44 (1).

Meehl, Paul E. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence.* Minneapolis, MN: University of Minnesota Press.

Milgram, Anne, Alexander M. Holsinger, Marie Vannostrand, Matthew W. Alsdorf. 2014. "Pretrial Risk Assessment: Improving Public Safety and Fairness in Pretrial Decision Making." *Federal Sentencing Reporter* 27 (4): 216–21.

Mohler, George O., Martin B. Short, Sean Malinowski, Mark Johnson, George E. Tita, Andrea L. Bertozzi, and P. Jeffrey Brantingham. 2015. "Randomized Controlled Field Trials of Predictive

Policing." *Journal of the American Statistical Association* 110 (512): 1399–1411.

Mullainathan, Sendhil, and Ziad Obermeyer. 2017. "Does Machine Learning Automate Moral Hazard and Error?" *American Economic Review* 107 (5): 476–80.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53.

Peterson, Richard R. 2020. *Brief No. 46: CJA's Updated Release Assessment.* New York: Criminal Justice Agency.

Rambachan, Ashesh, and Jonathan Roth. 2020. "Bias In, Bias Out? Evaluating the Folk Wisdom." Paper presented at 1st Symposium on the Foundations of Responsible Computing (FORC 2020), LIPIc, June 1–3.

Rambachan, Ashesh. 2021. "Identifying Prediction Mistakes in Observational Data." https://asheshrambachan.github.io/assets/files/asheshr_identifyingPredictionMistakes_Main.pdf.

Ridgeway, Greg. 2013. "The Pitfalls of Prediction." *NIJ Journal* 271: 34–40.

Ross, Stephen L., and John Yinger. 2002. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press.

Rudin, Cynthia, Caroline Wang, and Beau Coker. 2020. "The Age of Secrecy and Unfairness in Recidivism Prediction." *Harvard Data Science Review* 2 (1).

Rudin, Cynthia, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. 2021. "Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges." arXiv:2103.11251.

Ruhland, Ebony L. 2020. "Philosophies and Decision Making in Parole Board Members." *Prison Journal* 100 (5): 641–61.

Salzinger, Kurt. 2005. "Clinical, Statistical, and Broken-Leg Predictions." *Behavior and Philosophy* 33: 91–99.

Schuessler, Karl F. 1954. "Parole Prediction: Its History and Status." *Journal of Criminal Law, Criminology, and Police Science* 45 (4): 425–31.

Schwartzapfel, Beth. 2019. "Can Racist Algorithms Be Fixed?" *The Marshall Project*, July 1. https://www.themarshallproject.org/2019/07/01/can-racist-algorithms-be-fixed.

Smith, Aaron. 2018. "Public Attitudes Toward Computer Algorithms." *Pew Research Center*, November 16. https://www.pewresearch.org/internet/2018/11/16/attitudes-toward-algorithmic-decision-making.

Stevenson, Megan. 2018. "Assessing Risk Assessment in Action." *Minn. Law Review* 103 (303).

Stevenson, Megan T., and Jennifer L. Doleac. 2021. "Algorithmic Risk Assessment in the Hands of Humans." https://ssrn.com/abstract=3489440.

Tschandl, Philipp, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda et al.2020. "Human–Computer Collaboration for Skin Cancer Recognition." *Nature Medicine* 26: 1229–34.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.

Wang, Caroline, Bin Han, Bhrij Patel, Feroze Mohideen, and Cynthia Rudin. 2020. "In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction." *arXiv preprint arXiv:2005.04176.*

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 97–122

Inside the Box: Safety, Health, and Isolation in Prison

Bruce Western

Incarceration is part of the scholarly analysis of crime and inequality, but penal institutions largely remain a black box. Incarceration is often treated as a warehouse that removes people from society, or a diploma mill that confers the negative credential of a prison record. Unlike research on other institutional settings, like schools or hospitals, what happens inside prisons and jails does not figure greatly in the analysis of their effects. This essay will seek to look inside the black box and examine social dynamics inside prisons. What happens inside prisons and jails matters, because of the extreme scale of incarceration and the harsh conditions of penal confinement in the United States.

The term "mass incarceration" has come to refer to the exceptional scale of the US prison population (Garland 2001). Three empirical markers shown in Figure 1 describe mass incarceration. First, the rate of imprisonment in the United States increased from 93 per 100,000 in 1972 to its peak of 506 in 2007 (Panel A). Despite a decline in the last decade, the US imprisonment rate in 2018 was about four times higher than its general level during the twentieth century. Second, the US incarceration rate is about six times higher than is common in the nations of western Europe (Panel B). Indeed, the United States has the highest incarceration rate in the world (Walmsley 2019). Third, incarceration has become pervasive for Black men with no more than a high school education. Compare the cumulative risk of ever having been to prison up to age 35 for men born in 1945–1949 and three

■ *Bruce Western is the Bryce Professor of Sociology and Social Justice and Co-Director of the Justice Lab, both at Columbia University, New York City, New York. His email address is bruce.western@columbia.edu.*

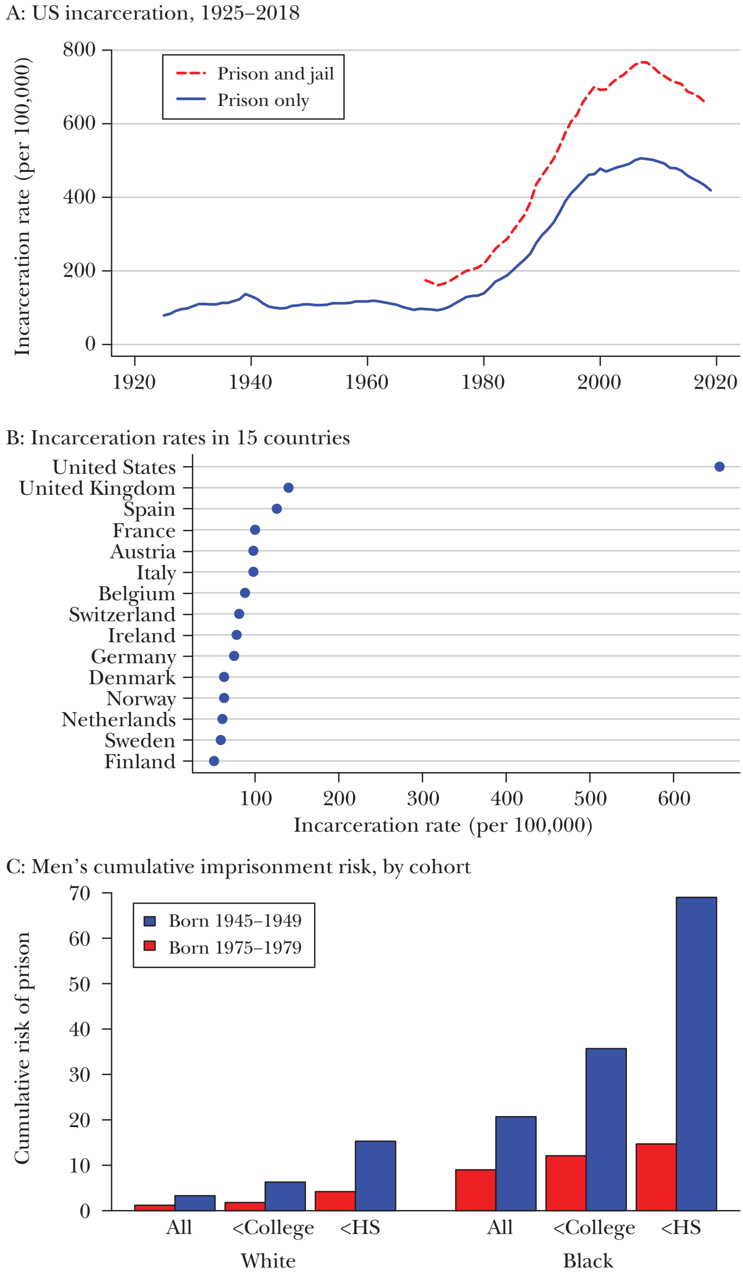
For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.97.

decades later from 1975 to 1979 (Panel C). The prevalence of incarceration is about five to seven times higher for Black men compared to Whites and is concentrated in the non-college fraction of the population. The prevalence of imprisonment also increased from the older generation to the younger: nearly 70 percent of Black men in the younger cohort who had not completed high school had been imprisoned by their mid-30s. The growth in incarceration resulted primarily from policy changes that imposed mandatory and longer prison sentences for a wide variety of offenses (Neal and Rick 2014; Travis, Western, and Redburn 2014; Raphael and Stoll 2013a).

Researchers have tried to understand the significance of mass incarceration by studying its effects. One line of work studied the effect of incarceration on crime: for reviews, useful starting points include Donohue (2009) and Durlauf and Nagin (2011). The crime-reducing effects of incarceration are often thought to operate through two channels: deterrence and incapacitation (for example, Bushway and Paternoster 2009). Incarceration deters those coming out of prison from recidivism, and the threat of prison deters would-be offenders. Incapacitation, on the other hand, removes people from society who would otherwise commit crime. Empirical studies of the effects of incarceration on crime have examined variation across states and micro-data on individual defendants, often exploiting changes in penal policy to isolate exogenous shifts in incarceration (Levitt 1996; Helland and Tabarrok 2007; Lofstrom and Raphael 2016). In the 1970s and 1980s, rising incarceration was associated with reductions in crime, but the crime-reducing effect of incarceration declined with scale. By the 1990s and 2000s, increases in a high level of incarceration had little effect on reducing violent crime. Much of the estimated effect since the 1970s appears to be due to incapacitation (Raphael and Stoll 2013a).

A second line of research examined the effects of incarceration on social and economic inequality, focusing on labor markets, health, and families and children (Neal and Rick 2014; Wakefield and Wildeman 2013; Western 2006; Pager 2003). In the labor market, incarceration is thought to reduce future employment and earnings by reducing work experience and creating the stigma of a prison record. Analysis of survey data on labor market effects often finds negative effects of incarceration (Grogger 1995; Apel and Sweeten 2010); however, null effects are often reported from natural experiments that exploit the random assignment of judges who vary in the severity of their sentences (Kling 2006; Harding et al. 2018). Despite mixed results, labor market research consistently finds that employers are reluctant to hire job seekers with criminal records, and levels of earnings after incarceration are extremely low (Pager 2003; Kling 2006; Dobbie, Goldin, and Yang 2018). Health studies report high risks of mortality immediately after release, related to drug overdose, as well as elevated levels of depression and other mood disorders (Binswanger et al. 2013; Schnittker, Massoglia, and Uggen 2012). Studies of families and children find that couples are more likely to divorce or separate during incarceration (Lopoo and Western 2005). Children whose fathers were incarcerated are at high risk of school suspension, health problems, homelessness, and perhaps criminal involvement (Turney and Goodsell 2018; Dobbie et al. 2019; cf. Norris, Pecenco, and Weaver 2021). Although causal inference is elusive in observational studies,

Figure 1
Mass Incarceration in Three Figures



Source: See Data Appendix
Note: Panel A shows the US imprisonment rate (1925–2018) and prison and jail incarceration rate (1972–2018); Panel B shows the incarceration rates in the United States and selected nations of western Europe; and Panel C shows the percentage of men who have ever been in prison by 1979 for those born 1945–1949, and by 2009 for those born 1975–1979, by race and education, including all men, those without college education, and those with less than twelve years of completed schooling.

incarceration has clearly become prevalent in low-income communities of color and is closely associated with poor adult health and life adversity among children.

This paper thinks about the significance of mass incarceration in a third way that focuses on the exceptionalism of US penal institutions. The sociologist Erving Goffman (1961, p. xiii) called the prison a "total institution," "a place of residence and work where a large number of like-situated individuals, cut off from the wider society . . . lead an enclosed, formally administered round of life." Prisons in this sense are like boarding schools, orphanages, or mental hospitals. However, prisons are for punishment, which provides few incentives to maintain or improve the welfare of incarcerated people. Incarcerated people also have few avenues within prisons or through the courts for self-protection against harmful conditions that go beyond the deprivation of liberty for which the law provides.

I examine four topics that shed light on the experience of imprisonment and help to illustrate the larger significance of mass incarceration. I first describe how increased incarceration accompanied prison overcrowding and reductions in rehabilitative programming. Next, I consider research on violence in prison and the safety of prisons compared to communities. I then discuss the health problems of incarcerated people and the risks of mortality and infectious disease in prisons. Finally, I examine the extreme isolation of solitary confinement.

Focusing on what happens inside prisons influences how we understand its effects, its costs, and its moral status. Violent victimization, impaired health, and the trauma of solitary confinement illuminate mechanisms by which incarceration may impair adjustment to community life. By examining the welfare of incarcerated people, we obtain a more complete accounting of costs and benefits, including the violent victimization of people in prison that is usually ignored in assessments of incarceration's effect on crime. Studying the harms that are suffered in incarceration raises the question of whether prisons are meeting a moral standard of dignified treatment and human respect.

The Retreat from Rehabilitation

US prisons of the early nineteenth century were conceived as being on the cutting edge of social reform, with the stated purpose being for the correction of people convicted of crimes (Rothman 2002a). Facing a post-colonial society undergoing rapid social change, social reformers of the Jacksonian period traced crime to the crumbling authority of the family and the church and an escalating depravity in social life. The prison sought to answer the moral decline that nourished crime with "doctrines of separation, obedience, and labor [that] became the trinity around which officials organized the penitentiary" (Rothman 2002a, p. 105). Work and strict discipline, often under conditions of silence and isolation, formed the routine of the original penitentiaries in New York and Pennsylvania of the 1820s. Conditions were harsh by modern standards, but the goal of rehabilitation was novel and more humane than the colonial penal code that prescribed whipping, stocks, or the gallows for a wide range of offenses. By the Progressive period of the early twentieth century, newly designed prisons resembled factories and schools, reflecting the central importance of employment and education to prison reformers (Rothman 2002b).

Rehabilitative effort was concentrated in the prisons of the Northeast and the Midwest, and the philosophy gained less influence in the South. There, the social organization of slavery was imprinted upon penal practice. The emergence of convict leasing and plantation-style prisons in the South following post–Civil War Reconstruction continued the earlier historic forms of forced labor and economic exploitation (Oshinsky 1997; Perkinson 2010; Muller 2018).

Work and education programs came to be widely used in American prisons, but the project of rehabilitation was caught in a struggle between what Rothman (2002b) called "conscience and convenience." Reformers hoped that rehabilitation would improve living standards and the quality of civic life. In practice, prison administrators had neither the policy knowledge nor the resources to help incarcerated people chart new life paths. By the post–World War II period, evaluations of rehabilitation programs often showed little success in reducing recidivism and improving welfare.

By the 1970s, a skepticism of rehabilitation converged with calls from the left to limit discretion in criminal justice decision-making (American Friends Service Committee 1971), and calls from the right to get tough on crime (Flamm 2005; Weaver 2007). In 1976, Robert Martinson and his co-authors published a comprehensive review of correctional programming that concluded that no particular kind of treatment consistently reduced recidivism (Lipton, Martinson, and Wilks 1975). Martinson went further in a *60 Minutes* interview a year later, saying that rehabilitation programs "simply have no fundamental effect on the recidivism rate of people who go through . . . the system" (as quoted in Cullen 2013, p. 327). Doubts about the effectiveness of rehabilitation were later echoed in a report by the National Research Council (Martin, Sechrest, and Redner 1981).

Statements of skepticism were soon answered with a defense of rehabilitation (Cullen 2004) and opinion among researchers began to shift once again during the 1990s. A modern theory of rehabilitation emerged that emphasized changing opportunities, behavior, and social bonds. Education and work programs, which had been staples of prison programming for a century, obtained support from a "life course" theory of crime in which stable employment provided structure and routine in daily life, and diverted workers from peers, who themselves may be involved in crime (Sampson and Laub 1990; Sampson and Laub 1993; Uggen 2000). Canadian researchers and practitioners developed a theory of rehabilitation that emphasized principles of risks, needs, and responsivity (Andrews, Bonta, and Hoge 1990; MacKenzie 2006, chapter 4). The risk principle argued that treatment should be proportional to the risk of recidivism; those who are most like to re-offend should receive the most intensive interventions. The needs principle says that programs should target known predictors of offending, including economic need and impulsive behavior. Finally, responsivity means targeting deficits that are susceptible to change. Age and gender, for example, are major risk factors for crime but they cannot be changed by rehabilitation. Researchers have found that anti-social attitudes and criminally involved peers are predictive of crime, and these have emerged as targets for case management and behavioral health programs.

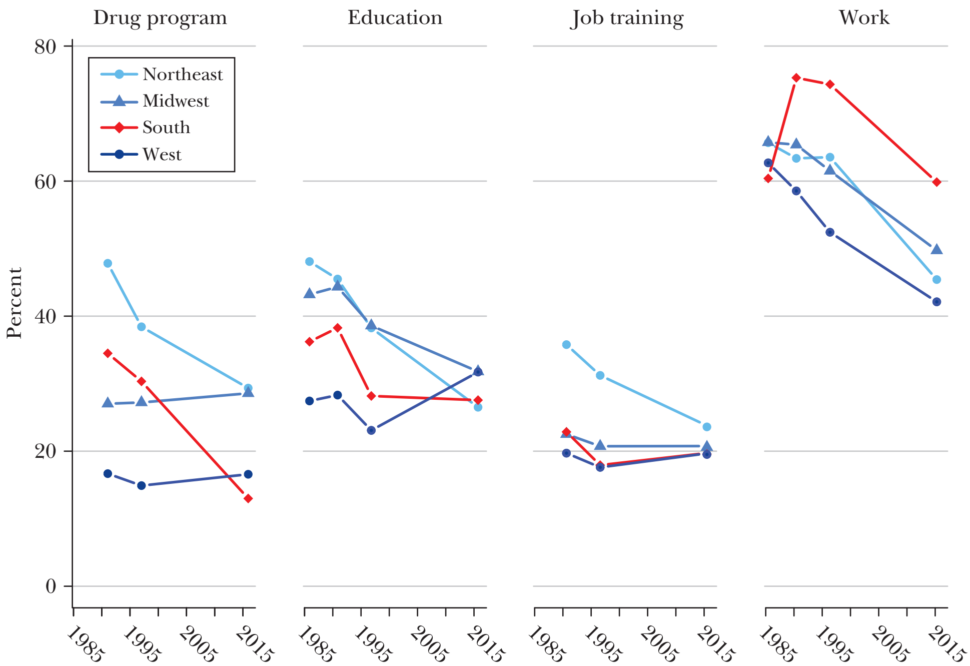
How strong is the evidence for rehabilitation programs? Gaes and his colleagues (2000, p. 361) summarized opinion among researchers: "Most correctional treatments for adult prisoners probably have modest positive effects." Stronger effects, they say, are observed for adolescents rather than adults, and community programs yield larger effects than programs in prison. Work and education programs are often found to be associated with reduced recidivism (Bozick et al. 2018; Gaes et al. 2000). For example, the Washington State Institute for Public Policy, which regularly surveys evaluations of correctional programs, reports that basic education, post-secondary education, and correctional industries all pass a cost-benefit test, and the effects are especially large for post-secondary education (Bitney et al. 2017).

Although research opinion rallied behind rehabilitation, the field offers few well-powered randomized trials. One of the few randomized experiments that includes educational programming in prison, the Milwaukee Safe Street Prisoner Release Initiative, provided a package of services that also included cognitive behavioral therapy, drug and alcohol treatment, case management, and post-incarceration support. One year after prison release, re-arrest rates for the 106 men in the treatment group were lower (63 versus 72 percent) than for the 130 men in the control group, but employment outcomes were similar for treatment and control subjects (Cook et al. 2015). Behavioral interventions, such as cognitive behavioral therapy or motivational interviewing, are also often found to be associated with reduced recidivism. The CrimeSolutions website of the US Department of Justice, for example, classifies both kinds of interventions as "Promising," meaning they are supported by "moderate quality evidence with statistically significant average effect sizes" (Office of Justice Programs 2021).

In an area of evaluation that often struggles with program fidelity and research design, the case for rehabilitation has been buttressed by meta-analysis that pools together large numbers of (often imperfectly designed) studies. Meta-analysis consistently finds that deterrence is less effective than rehabilitation at reducing recidivism: "Interventions that are punitive—that emphasize deterrence, discipline, or surveillance—have weak . . . effects on recidivism" (Cullen 2017, p. 248).

Despite the shifting weight of the evidence, policymakers of the 1980s and 1990s largely rejected rehabilitation and adopted incapacitation and deterrence as the main goals of penal policy. The Survey of Inmates of State Correctional Facilities (renamed the Survey of Prison Inmates in 2016) periodically interviews respondents in state prison. Figure 2 shows participation rates in drug programs, education, job training, and work assignments reported in the survey from 1986 to 2016 across regions of the country. Program participation tends to be highest in the Northeast and lower in the South and the West. The figure shows broad reductions in program participation in US prisons across the country and across program areas. Participation in drug treatment programs in the Northeast and the Midwest fell from highs of 30 to 50 percent in the 1980s to below 20 percent by

Figure 2
Enrollment in Drug, Education, Job Training Programs, and Work Assignment, State Prisoners, by Region



Source: Authors calculations from the Survey of Inmates of States Correctional Facilities (1986, 1991, 1997, 2004) and the Survey of Prison Inmates (2016) (Bureau of Justice Statistics 1994, 2004a, 2004b, 2006, 2019, 2021).

2016. Educational programming was also reduced in all regions except the West. Job training became less common. Participation in work assignments also fell across the country. Although participation could be explained by the availability of programs or enrollment given availability, the trends are consistent with other evidence of reduced opportunities for academic education and work release employment in American prisons (Phelps 2011; Jung 2014, p. 385).

Declines in program participation and the turn away from rehabilitation had implications for life inside prisons. Rehabilitation has a symbolic component, signaling society's commitment to compassion and an individual's capacity for change. Rehabilitation announces that we are "a civilized nation . . . we are capable of turning our collective cheek in hopes of effecting redemption" (Cullen 2013, p. 308).

The retreat from rehabilitation was accompanied by other changes in the internal dynamics of penal institutions. As the goals of imprisonment shifted to incapacitation and deterrence, prison populations grew rapidly in the 1980s and 1990s and overcrowding added to the harshness of prison conditions. Overcrowding has been a longstanding focal point of litigation for US prisons (Levitt 1996; Simon 2014; Guetzkow and Schoon 2015). A series of lawsuits in California revealed the effects of overcrowding on the delivery of health care and prison safety. At its peak in 2007, California accounted for 13 percent of all state-level prisoners nationwide and the system had regularly operated at 150 percent or more of its designed capacity at least since the 1980s. California prisons were so overcrowded in the early 2000s that gymnasiums were converted into housing units and triple-bunking was used in some facilities. To describe the conditions of incarceration, Simon (2014, p. 117) quoted from a federal court opinion following Governor Arnold Schwarzenegger's State of Emergency proclamation on prison overcrowding in 2006:

> The risks enumerated by the Governor in his Proclamation include "increased, substantial risk for transmission of infectious illness"; security risks caused by line-of-sight problems for correctional officers, particularly in areas where inmates are triple-bunked and in "tight quarters"; and "thousands of gallons of sewage spills and environmental contamination" from overloading the prisons' sewage and wastewater systems . . . . Governor Schwarzenegger also declared that the suicide rate in the 29 severely overcrowded prisons "[was] approaching an average of one per week."

Health care failures due to overcrowding ultimately caused a panel of federal judges to rule that the entire California state prison system of more than 150,000 incarcerated people was unconstitutional, in violation of the Eighth Amendment prohibition against cruel and unusual punishment.

Prison releases in the context of overcrowding litigation have been used by researchers as sources of exogenous variation that can identify incarceration's effect on crime (Levitt 1996; Lofstrom and Raphael 2016; Sundt, Salisbury, and Harmon 2016). Our current focus on prison conditions raises the question of whether overcrowding may have harmed the people who were released. Because health care, programming, and safety are all diminished by overcrowding, the effect of releases linked to overcrowding lawsuits may be different from, say, the effects of changes in incarceration produced by sentencing reform. The harm suffered by men and women in overcrowded California prisons may impair adjustment after incarceration and perhaps ultimately threaten public safety. The empirical evidence shows that crime did not increase greatly following court-ordered releases in California (Lofstrom and Raphael 2016), suggesting that the while the post-rehabilitative prison may be seriously harmful, it is not necessarily criminogenic.

Safety and Victimization

In the period of rapid growth in incarceration when overcrowding became persistent for many states, researchers often described prisons as warehouses, operating chiefly as storage units for prime-age men of color from poor communities

(Lynch 2009; Phelps 2011). The loss of programs and ensuing idleness, overcrowding, and the influx of new and younger prisoners have all been associated with violence in prison. How violent are prisons, and have they become more violent as the incarceration rate has grown?

Violence and the fear of violence casts a broad shadow over prison life. Violent victimization in prison includes threats, assaults, rapes and other sexual assaults, and homicide (Bowker 1980). The classic prison field study, *The Society of Captives* by Gresham Sykes (1958), described the loss of security as one of five "pains of imprisonment."[1] The penologist Hans Toch (1977) found safety of the prison environment to be among the most important concerns of incarcerated people. Many studies using a wide variety of data consistently find higher levels of violence in prison than in the general population (Bottoms 1999). As a form of violent self-harm, suicide rates are also higher in incarceration than in the general population. Alison Liebling (1999, p. 341) writes that "fear, anxiety, loneliness, trauma, depression, injustice, powerlessness, violence, rejection, and uncertainty are part of the experience of prison" in which "suicide is perhaps its most dramatic outcome." Violent victimization and a fear of violence can lead to social withdrawal, hypervigilance, a tough exterior, and flat affect as incarcerated people try to avoid conflict (Haney 2006, pp. 172–73). Victimization during incarceration is also associated with later drug use, emotional distress, depression, and criminal offending (Wooldredge 1999; Zweig et al. 2015; Hochstetler, Murphy, and Simons 2004).

Violence in prison has been measured both by official statistics and by surveys of self-reported offending and victimization. Prison rules prohibit assaults, fighting, violent threats, and the possession of weapons. Prison staff can bring charges and write tickets for misconduct, similar to police in community settings. Like community arrest records, prison records on misconduct underestimate violence. Violence often goes undetected by staff, incarcerated people are reluctant to report violence to authorities, and staff have wide discretion in responding to the violence they do encounter. Self-report data in prison indicate that violent victimization and patterns of repeated victimization are more common than official infractions (Cooley 1993; Bottoms 1999, p. 222). Still, self-reports in prison may also underestimate victimization. In my own fieldwork, formerly incarcerated survey respondents spoke more readily about violence in prison after they were released than while incarcerated. Where victimization elicits shame or embarrassment, as it may in the congressionally mandated surveys of prison rape and other sexual violence, underreporting is also likely.

One indication of the scale of violence in US prisons is provided by data on homicide victimization. Similar to community trends, the prison homicide rate has declined significantly from the 1980s to the late 2010s. In 1980, the homicide rate in state prison was 54 per 100,000 compared to 10.2 per 100,000 in the US population.

[1] The four others were the loss of liberty, desirable goods and services, heterosexual relationships, and autonomy.

The prison homicide rate fell steadily to a rate of 3 per 100,000 in 2001 but has since increased to 8 per 100,000 in 2016 (Mumola 2005; Carson and Cowhig 2020).

Because offending and victimization varies by age, race, and gender, the relative risk can be obtained by comparing men's age-specific homicide rates in prison to general-population homicide rates calculated from Vital Statistics, adjusting for racial composition. The adjustment takes an average of race-specific homicide rates weighted in proportion to the racial composition of the prison population.[2] Adjusting for racial composition accounts for the over-representation of Black and Latino men in prison whose victimization rates in the general population are relatively high. As shown in Table 1, men under 35 years old face significantly lower rates of homicide victimization in prison than in the general population (Table 1). Unusually compared to community patterns, the prison victimization risk increases with age over age 45 (about one-third of the prison population). For older prisoners over age 55, the homicide victimization rate is equal to the rate in free society.

There is less lethal violence in incarceration than in the general population in part because of the absence of firearms in prison. Correctional officers working their usual shifts in housing units, dining halls, and recreation areas do not carry guns and gun deaths in prison are rare. In the community, on the other hand, guns account for nearly 90 percent of all homicides. We can adjust for the lethality of firearms by re-calculating population homicide rates, counting only non-firearm deaths. The homicide rate in prison is two to three times higher than the non-firearm homicide rate in the general population, reflecting the high level of manual violence in prison. The total age-race-adjusted rate for non-firearm homicides in the community is about half the prison homicide rate (as shown in Table 1, 3.5 in the population compared to 6.6 per 100,000 in prison).

The last two columns of Table 1 examine non-lethal violence in prisons by comparing prison infractions to self-reported victimization in the community. Data on prison infractions are taken from the Survey of Prison Inmates (2016), in which respondents were asked if they were written up for assaulting an officer or another inmate in the last 12 months. Figures on violent victimization in the community were calculated from the National Crime Victimization Survey (2016) that asked respondents whether they had been attacked or threatened in the last 12 months (Bureau of Justice Statistics 2020). Similar to the homicide analysis, the community victimization rates are weighted by the racial composition of the prison population. Prison infractions and victimization rates measure different things, but both indicate levels of violence. The prison infraction rate certainly underestimates the level of violent offending. The community victimization rate is a more direct measure of harm and includes incidents that are unreported to the authorities. The infraction

[2] For age groups $a = 18–24, 25–34, \ldots, 55–85$ and race groups $r = B, W, L$ (for Blacks, Whites, and Latinos), we have prison population shares p_a and p_r, and race-age-specific homicide rates in prison, h^P_{ra}, and in the general community, h^C_{ra}. At age a, the prison race-adjusted homicide rate is given by $h^P_a = \sum_r p_r h^P_{ra}$ and the community race-adjusted homicide rate is $h^C_a = \sum_r p_r h^C_{ra}$, reflecting the race distribution of the prison population. The total age-adjusted homicide rate in the community is given by $h^C_a = \sum_p p_a h^C_a$, reflecting the age distribution in prison.

Table 1
Violence and Victimization in Prison

Age group	Homicide victimization rate: Deaths in prison	Homicide victimization rate: All deaths in the population	Homicide victimization rate: Non-firearm deaths in the the population	Assault infractions in prison	Violent victimization in the population
Total	6.6	26.4	3.5	56.4	9.9
18–24	6.1	35.7	2.6	128.6	13.9
25–34	5.5	37.5	3.5	74.9	10.6
35–44	6.9	24.9	4.2	45.2	8.5
45–54	6.9	12.6	3.1	29.5	8.7
55 or older	8.7	8.5	3.8	20.4	6.2

Source: Author's calculations. Prison homicide rates are estimated from BJS (2020) and population homicide rates are calculated from CDC vital statistics (Carson and Cowhig 2020). Infraction rates for assault in prison are estimated from the Survey of Prison Inmates (2016) (Bureau of Justice Statistics 2021). Violent victimization rates are estimated with the National Crime Victimization Survey (2016) (Bureau of Justics Statistics 2020).
Note: Men's rates of homicide victimization are expressed per 100,000. Enforcement actions for assault infractions in prison and violent victimization in the general population are expressed per 1,000. Prison homicide victimization rates are for 2012–2016 and population homicide rates are for 2016. Population rates are adjusted by weighting age-race-specific rates by the age and racial composition of the prison population. See text for details.

rate and the victimization rate are both age-graded. Whether in prison or not, young men are most likely to be involved in violence, either as offenders or victims. The prison infraction rate is three to eight times higher than self-reported victimization in the community crime survey. The overall age-adjusted rate for violent infractions in prison is more than five times higher than the community-based rate of violent victimization. The data indicate that assaults and fighting are significantly more common in prison than in free society, accounting for the age and racial composition of the prison population.

In a context of rising imprisonment rates through the 1980s and 1990s, advocacy organizations became more concerned about violence in prison. "Overcrowded and understaffed, filled with too many idle prisoners facing long terms of incarceration, many U.S. penal facilities are rife with extortion, violence, and other abuses," wrote Human Rights Watch in 2001 (Mariner 2001). Their report called out the casual acceptance of rape in men's prisons on the part of correctional administrators. In 2003, Congress passed the Prison Rape Elimination Act, which started a regular data collection on the prevalence of sexual violence and initiated new anti-violence protocols for American prisons. The Bureau of Justice Statistics now fields annual surveys, but the true prevalence of sexual violence in prison is difficult to estimate. Prevalence estimates vary widely from about 1 to 20 percent depending on the reference period and the survey methods. For example, a self-administered questionnaire sent to seven prisons in the Midwest in the mid-1990s yielded a rate of coerced sexual activity during the current incarceration of 210 per 1,000 (Struckman-Johnson and Struckman-Johnson 2000). In the early 2000s, survey estimates of the six-month prevalence of nonconsensual sex acts ranged from around 15 to 30 per 1,000 (Wolff et al. 2006). Illustrating the difficulties of measuring sexual violence in prison, self-reported victimizations in a prison survey jumped from 10,000 in 2012 to nearly 25,000 in 2015 after protocols were introduced for investigating allegations of sexual violence (Rantala 2018). Despite the great variability, even the low-prevalence estimates in prison exceed the community estimates of rape or sexual assault where the 12-month prevalence is 1 to 2 per 1,000 (Morgan and Truman 2020).

Research on violence in prison has implications for how we understand the crime-reducing effects of incarceration. Studies of incapacitation and deterrence rely on community-based measures of crime that take no account of violent victimization in prisons. Incapacitation studies treat prisons as crime-free, but the levels of assault and sexual violence in prison are significantly higher than in the community. High rates of violence inside prisons, where punishment is certainly more swift if not more certain than in the community, also appears to be inconsistent with the idea that people will be deterred from misconduct if faced with immediate disciplinary action.

Health and Health Care

As prisons grappled with escalating populations and new legislative scrutiny on problems of violence, the costs of health care in prison were climbing. California, the leader on correctional health expenditures, spent an average of \$21,847 on medical expenses per person incarcerated in state prison in 2015, compared to \$7,807 in 2001 (in 2021 dollars) (Schiff et al. 2014; Huh et al. 2017). The growth in health care costs is related to the rising real cost of health care in the economy as a whole, and also to the aging of the prison population that accompanied the increasing length of prison sentences.

Even without population aging, incarcerated people are generally in poor health and have high needs for health care. Epidemiological data indicate three areas in which incarcerated people are in worse health than the general population: chronic conditions, infectious disease, and mental illness. People in prison suffer from poor physical and mental health that is often related to persistent poverty over the life course and risky health behaviors such as needle use, heavy alcohol use, and smoking (Fazel and Baillargeon 2011). Rates of chronic conditions like hypertension, asthma, and arthritis are about 50 percent higher in prison than in the community (Binswanger, Krueger, and Steiner 2009; Fazel and Baillargeon 2011). The prevalence of serious mental illness such as bipolar disorder and psychotic conditions like schizophrenia are around five times higher in prison than in the general population (Raphael and Stoll 2013b).

Despite the relatively high burden of disease, mortality rates in prison are not uniformly high. Standardized mortality ratios for White men have been estimated at around 1.2, indicating an age-standardized mortality risk in prison about 20 percent higher than in the general population (Rosen, Wohl, and Schoenbach 2011; Patterson

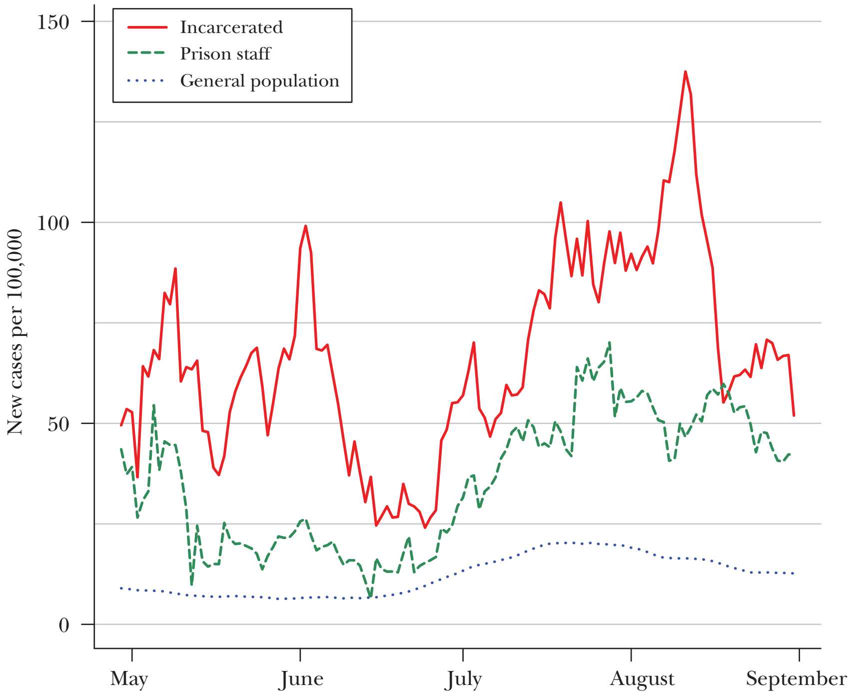
2010). Conversely, standardized mortality ratios estimated for Black incarcerated men have been estimated at around 0.5, indicating the death rate for Black men in prison that is about half the death rate for those in the general population (see also Wildeman et al. 2016). Patterson (2010) examined the contribution of violence to prison and community mortality rates and found that the low prison homicide rate of Black men could not explain the mortality gap between prison and community. Analysis of cause-specific mortality data for men incarcerated in North Carolina found that the excess risk was associated with cardiovascular disease, cancer, and infectious disease (Rosen, Wohl, and Schoenbach 2011). While few studies directly the examine the causes of low mortality among incarcerated Black men, researchers speculate that correctional health care improves the everyday treatment of chronic conditions compared to the quality of health care in free society (Rosen, Wohl, and Schoenbach 2011; Patterson 2010). Other characteristics of prison life such as regular meals and consistent housing may also help the management of chronic conditions.

Although correctional healthcare may reduce mortality for Black men with chronic conditions, prison clearly impairs health along other dimensions by elevating the transmission of infectious disease. High rates of HIV and hepatitis B and C have been widely documented in US prisons. Recent estimates indicate that HIV prevalence in prison exceeds community rates by a factor of 3 to 5, and the prevalence of hepatitis B and C exceeds community rates by 5 to 10 times (Bick 2007; Gough et al. 2010). Screening at prison intake suggests around 80 to 90 percent of cases were present before incarceration, with the remainder transmitted in prison, mostly through sexual activity and needle use. A related line of research studies outbreaks of infectious disease, focusing on tuberculosis, influenza, and chickenpox (Beaudry et al. 2020). Each of these infections are airborne and spread through aerosol transmission (droplets) and contact with surfaces. The living areas, dining halls, and recreation areas that make up the physical plant of prisons facilitate the spread of airborne pathogens, particularly in overcrowded conditions.

The significance of correctional facilities for the transmission of infectious disease were strikingly illustrated by the novel coronavirus pandemic. Prisons and jails were consistently among the leading hotspots for COVID-19 outbreaks throughout 2020 (Wang et al. 2020). Facilities such as Rikers Island jail in New York City, Cook County Jail in Chicago, and Marion Correctional Institution in Ohio suffered ferocious outbreaks that resulted in dozens of fatalities among staff and incarcerated people.

Measuring the true prevalence of COVID-19 has been challenging because testing varies greatly across the population and across prison systems. Moreover, infection dynamics are highly nonlinear, so prevalence estimates are sensitive to time and space. One of the best case studies of COVID-19 dynamics estimated the reproduction number, called R_t in epidemiological models, in an unnamed county jail (Puglisi et al. 2021). The reproduction number quantifies the new infections associated with a single infected case. An $R_t > 1.0$ indicates an outbreak where the prevalence of infection increases nonlinearly. Testing for the novel coronavirus in the study jail included daily measures for 83 days in 2020. At the onset of

Figure 3
Seven-Day Moving Average of New Daily COVID-19 Cases per 100,000 Among Those in Prison, Prison Staff, and in the General Population: May to September 2020

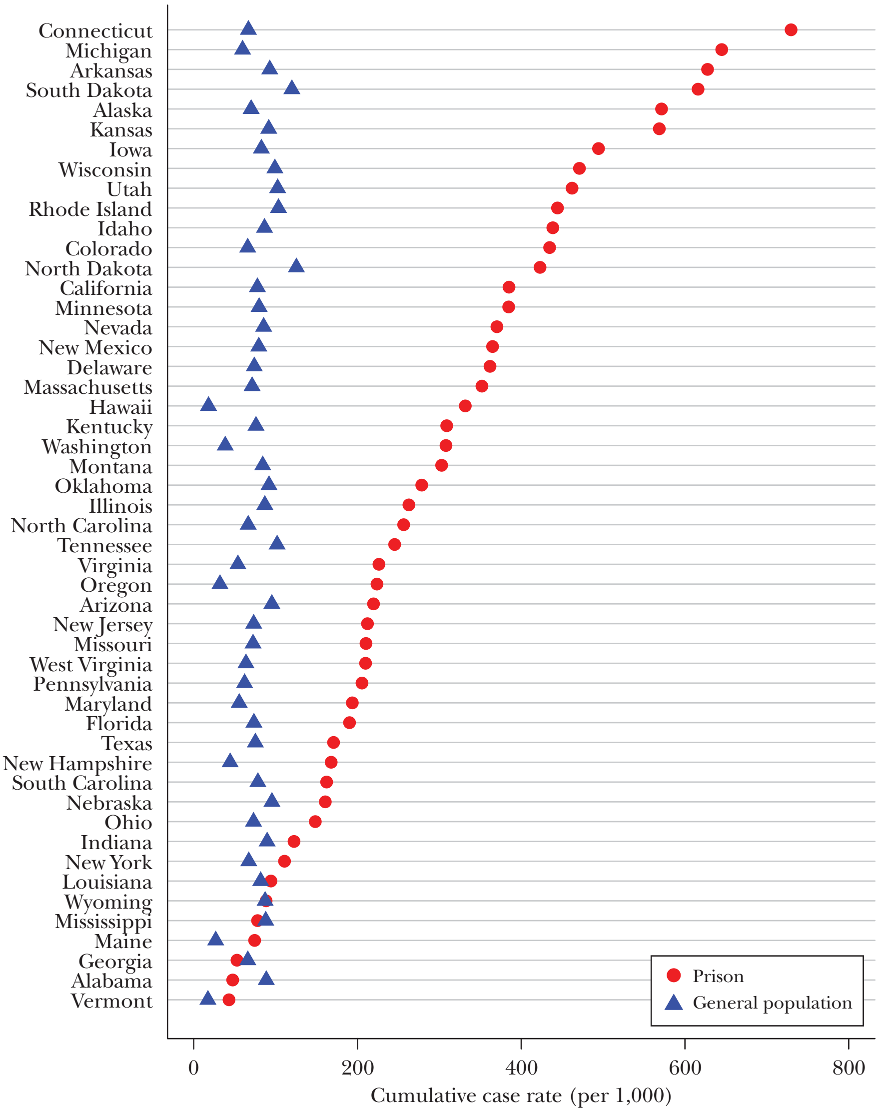


Source: Author's calculations from the COVID Prison Project file.

the outbreak in the jail, R_0 was estimated at 8.23, meaning that a single person with the novel coronavirus infected eight others. At this reproduction rate, the spread of infection in the jail population was explosive (Puglisi et al. 2021). Overcrowding accelerated the spread of infection. When the jail depopulated, R_t dropped below 1.0.

Obtaining a national picture of the scale of COVID-19 in American prisons has been more difficult than in the jail study. National data on case rates points to the high level of infection in incarceration. Figure 3 shows the daily incidence (seven-day moving average) of new COVID-19 cases expressed as a rate per 100,000 of the population, from May to September 2020. In the general US population, the daily incidence of new COVID-19 cases increased from 9 per 100,000 in May to a peak of 20 per 100,000 in late July. Among prison staff, the daily incidence of new cases averaged 36 per 100,000. Among incarcerated people, the daily incidence rate averaged 67 per 100,000 and peaked at 138 per 100,000 by August 2020. COVID-19 case rates in prison varied greatly across states, partly because of the pattern of outbreaks and partly as a function of measurement that depends on the level of testing. Still the measured COVID-19 case rate in prison exceeded the case rate in the general population in nearly all states, as shown in Figure 4.

Figure 4
Cumulative COVID-19 Case Rates among Those in Prison and General Population, by State: March 2020 to January 2021



Source: Author's calculations from the COVID Prison Project file.

Research on health and incarceration underscores the vulnerability of people in prison. The high rate of chronic conditions indicates health vulnerability. Evidence that prison protects against mortality for Black men reflects the health risks and inadequacy of care in their home communities. While prison is a venue that brings together large numbers in poor health, the physical organization of prison and the routine of daily life creates an environment that accelerates the transmission of infection.

Solitary Confinement

Poor health marks vulnerability to harsh prison conditions, and solitary confinement is a vivid marker of harsh conditions. In its official purposes, solitary confinement is used to punish misconduct, to control conflicts such as gang rivalries, and for the protective custody of those who are unsafe in the general prison population because of, say, youth or gender identity (Kapoor and Trestman 2016, p. 200). The uses of solitary confinement are reflected in vernacular that distinguishes "disciplinary segregation" for misconduct from "administrative segregation" for managing safety. In practice, the conditions of confinement often do not vary much between the punitive and administrative functions. Incarcerated people are generally locked in their cells for 22 or 23 hours each day with an hour out for recreation or showers. Usually people incarcerated in solitary confinement are restricted from having visits, phone calls, or participating in programs. Cloud and his co-authors (2015, p. 19) describe the physical space of a solitary confinement unit:

> The typical cell is 60 to 80 square feet, with a cot, a toilet, a sink, a narrow slit for a window, and sometimes a small molded desk bolted to the wall. In many facilities, cells have a steel door with a small slot for delivering meals . . . Some solitary confinement units are nearly silent except for sudden outbursts; others subject prisoners to incessant cacophony of clanking metal doors, jingling keys, booted footsteps, and distressed voices reverberating off thick walls.

Nearly all US prisons have housing units for disciplinary and administrative segregation. A census of prison facilities shows that the population incarcerated in solitary confinement increased from 3.0 percent of the total prison population in 1979 to 5.7 percent in 2005. The use of solitary confinement also expanded with proliferation of super-maximum-security prisons in the 1980s and 1990s (Reiter 2016). Supermax prisons house their entire populations under conditions of 23-hour lockdown, which is characteristic of solitary confinement. In the most recent data from 2016, 4.4 percent of a national sample of the US prison population was held in solitary confinement (Beck 2015). There are no detailed national statistics on the length of stay, but a survey of state correctional leaders found that 11 out of 24 jurisdictions held most incarcerated people in solitary confinement for 90 days or less (Liman Program and ASCA 2015).

Psychologists find clear evidence of mental distress in solitary confinement (Kapoor and Trestman 2016; Arrigo and Bullock 2008). A clinical psychologist, Stuart Grassian (1983), coined the term "SHU syndrome," named for the Special Housing Units of Massachusetts prison system. Clinical assessments of people housed in solitary confinement revealed evidence of being in a mental fog, obsessive thoughts, perceptual distortions, hallucinations, and other forms of distress. Evidence for the negative effects of solitary confinement on mental health is especially strong for those who are in strict isolation for long periods and for those with a history of mental illness (Arrigo and Bullock 2008).

Mental illness and harsh conditions of penal confinement are intimately connected. Prisons bring together men and women with significant physical, mental, and behavioral health problems in a physical space that is often overcrowded, subject to a rule-governed climate, and managed by staff with broad discretion. People with mental illness may fail to respond to orders or have difficulty following prison rules. In some cases, they may act violently. Prison officials then respond to prisoners with mental illness "as they do to other prisoners who break the rules. When lesser sanctions do not curb the behavior, they isolate the prisoners in the segregation units, despite the likely mental health impact" (Metzner and Fellner 2013, p. 317).

We can look in greater detail at the population dynamics of solitary confinement and the overrepresentation of people with mental illness by analyzing an administrative data file from Pennsylvania that includes all prison admissions and discharges from 2008 to 2017. Pennsylvania's incarceration rate is approximately equal to the national level and the state's prison population is demographically similar to the national prison population as a whole. Like a number of states, the Pennsylvania prison system faced the threat of federal oversight for placing people with serious mental illness in solitary confinement. The state introduced a screening assessment that classifies all new admissions into one of four categories: 1) no mental health problems, 2) prior diagnosis of mental illness, 3) current diagnosis of a non-serious mental health problem, and 4) serious mental illness that includes psychotic conditions such as schizophrenia.

Table 2 reports figures describing the use of solitary confinement including disciplinary and administrative segregation in Pennsylvania, in the period 2008–2017. The median length of stay in prison for completed prison terms is about 18 months for men and 13 months for women. (This calculation underestimates length of stay in the whole prison population because long sentences are censored.) Solitary confinement is used more often for men than women. Although the national solitary confinement rate has been estimated at about 4 percent, around 40 percent of men and 25 percent of women in Pennsylvania are incarcerated in solitary confinement at some point during their imprisonment. Men are also more likely to be held repeatedly in solitary confinement. The data also indicate large differences in solitary confinement by mental health status. Among men with serious mental illness, 51 percent are in solitary confinement at some point during their prison sentence, compared to 32 percent for men with no mental illness. Those with serious mental illness are also repeatedly incarcerated in solitary confinement. Men with no history of mental illness who are ever sent to solitary confinement spend 37 days on average in isolation; in contrast, men with serious mental illness accumulate a total of 55 days on average in solitary confinement.

Mental health disparities are illustrated in greater detail by density plots of the total days in solitary confinement for those who have been held in isolation for

Table 2
Imprisonment and Solitary Confinement in Pennsylvania State Prisons, by Mental Health Status and Gender: 2008–2017

	All	*No mental illness*	*Prior diagnosis*	*Current diagnosis*	*Serious mental illness*
Men					
Median years of imprisonment	1.52	1.45	1.63	1.58	1.54
Proportion ever in solitary	0.39	0.32	0.42	0.49	0.51
Times in solitary (mean)	2.01	1.60	1.95	2.67	2.51
Median solitary spell (days)	28.50	24.00	27.00	27.00	22.00
Sample size (N)	162,763	82,542	42,854	33,782	3,675
Women					
Median years of imprisonment	1.09	1.02	1.21	1.11	0.98
Proportion ever in solitary	0.25	0.16	0.22	0.28	0.33
Times in solitary (mean)	1.80	1.22	1.41	1.94	2.16
Median solitary spell (days)	27.00	27.00	28.50	27.00	29.00
Sample size (N)	15,793	2,848	2,967	8,133	1,845

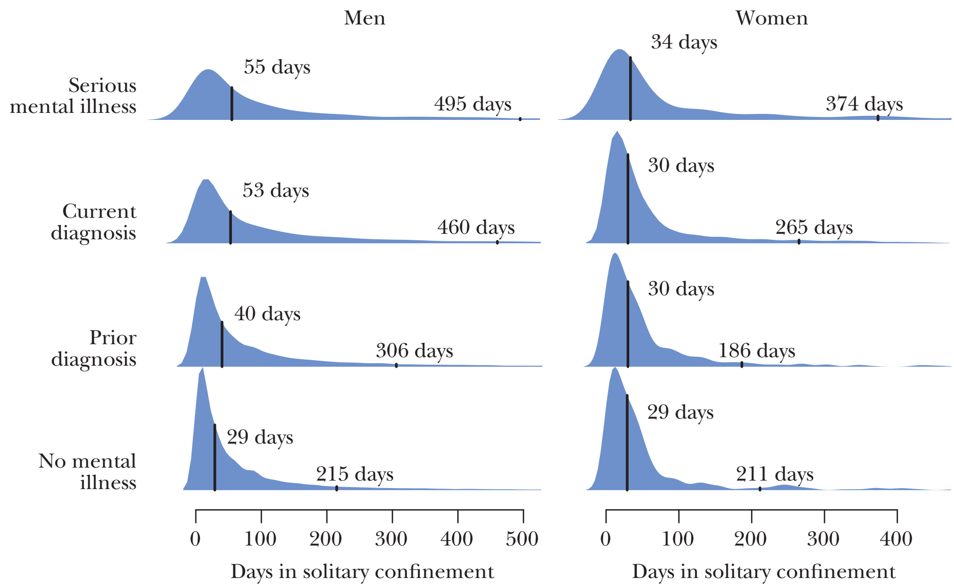
Source: Author's calculations with data provided by the Pennsylvania Department of Corrections.

at least one day. The total duration of solitary confinement for this Pennsylvania data is reported by mental health status in Figure 5. The median total duration for those who have ever been in solitary confinement is 29 days for men with no mental illness, compared to 55 days for those with serious mental illness. A small number of incarcerated people spend very long periods in extreme isolation. Among men with no reported mental illness, 10 percent have spent at least 215 days in solitary confinement. The mental health gap in the total time spent in solitary confinement at the 90th percentile is about nine months. This means that 10 percent of those men classified with serious mental illness spend 280 days longer in isolation than the top 10 percent of men with no mental illness. Women show a similar pattern, but they spend less time in solitary confinement and mental health disparities are smaller.

Solitary confinement has been scrutinized for its harmful effects and its ethical standing. Perhaps resulting from the injuries to physical and mental health, solitary confinement is associated with subsequent unemployment, recidivism, and mortality after prison release (Mears and Bales 2009; Brinkley-Rubinestein et al. 2019; Wildeman and Andersen 2020), although nonrandom selection of those with preexisting mental illness may also explain poor outcomes after prison release. Beyond measurable harm, solitary confinement has also raised questions of ethical treatment. Federal courts have examined solitary confinement in relation to the Eighth Amendment prohibition against cruel and unusual punishment and held that conditions must be compatible with "civilized standards, humanity, and decency" (*Madrid v. Gomez*, 889 F. Supp. 1146, 1260 [1995]). US health organizations have argued for limiting or abolishing solitary confinement because of

Figure 5

Density Plots of the Total Time Spent in Solitary Confinement by Mental Health Status and Gender, Pennsylvania State Prisons: 2008–2017



Source: Author's calculations from data provided by the Pennsylvania Department of Corrections
Note: Lines on the density plot indicate the median and 90th percentile.

its harm to physical and mental health (American Psychiatric Association 2012; National Commission on Correctional Health Care 2016). The so-called "Mandela Rules" established by the United Nations regards prolonged solitary confinement as a type of torture (United Nations General Assembly 2015). Under the inherently coercive circumstances of incarceration, the pains suffered in extreme isolation take on a moral significance.

Discussion

Human needs are often directly met through the intimate relations of households (the "oikos" in Greek, from which the word economics is derived). People eat, sleep, and shelter in dwellings with others to whom they are intimately tied through the connections of kin and clan (Weber 1978, pp. 356–381; Polanyi 1957, pp. 53–55; Goffman 1961, p. 12). Households woven together by personal relationships are good at meeting the variety of individual needs. But in total institutions, the satisfaction of basic needs is organized bureaucratically and uniformity is imposed on daily life. In prisons, uniformity is implemented through the power relations that divide staff from incarcerated people. Even in well-run prisons, the environment risks severe treatment and physical harm. With over half a million people entering prison each year, and a similar number returning to their communities, the footprint of institutional harm extends beyond the 1.4 million people who comprise the annual prison population.

Evidence of harsh prison conditions is connected to the emergence of mass incarceration and a distinctively American way of doing incarceration. The severity of US prison conditions contrasts with incarceration in Western Europe. One window into European prison conditions is provided by visits taken over the last ten years by US prison administrators to Germany, the Netherlands, and Scandinavia. European prisons were found to be organized around principles of "normalization" and "re-socialization" that aimed to close the gap between institutional conditions and conditions of life in free society (Subramanian and Shames 2013; Delaney et al. 2018). In Europe, terms of incarceration are significantly shorter and incarcerated people often obtain furloughs to visit family and work in the open labor market. High levels of security, solitary confinement, and prison uniforms that are characteristic of US incarceration are less common in Europe. Just as European prisons may be more humane, they may also be more positive in their effects. For example, research from Norway and Sweden points to the positive effects of incarceration on employment, earnings, and health after release (Bhuller et al. 2020; Hjarlmarsson and Lindquist 2020).

Harms suffered in US prisons suggest channels through which incarceration may diminish life chances. In the course of a prison term, our best evidence indicates that one in five incarcerated people may be physically or sexually assaulted, two in five may go to solitary confinement, and one in ten may acquire tuberculosis, hepatitis, or other infectious disease. Such harms transform incarceration from an expected cost of offending into an active influence on well-being after release. They may contribute to relapse to addiction, unemployment, or family estrangement that has been well-documented by researchers. Harmful experiences of incarceration may also help explain the criminogenic effects that result in offending after imprisonment (Chen and Shapiro 2007; Aizer and Doyle 2015; Nagin, Cullen, and Jonson 2009).

A focus on prison conditions should also expand our assessment of costs and benefits of mass incarceration. Research on the effects of incarceration on crime has underestimated the costs of incarceration by overlooking the risks of violent victimization in penal institutions. The neglect of fear and victimization inside prison is a flaw in studies of incapacitation and black-box estimates of the total effects of incarceration on crime.

While this paper has presented evidence of the harms suffered during incarceration, it does not answer the causal question of whether prisons are harmful. Compared to what? One counterfactual focuses on the conditions of life in the communities from which the prison population is drawn. For example, mortality risk is lower in prison for Black men, but higher for Whites. The risk of infectious disease, however, often appears to be higher in prison, and correctional facilities were hotspots for COVID-19. There are indications that violence—assaults and fighting, but not homicide—is more prevalent in prisons than community. The extreme isolation of solitary confinement seems deeply incomparable to community life. On some dimensions, it seems, prison may be no worse than the counterfactual conditions of community life, but on other dimensions, prison life is unusually painful.

The comparison of prison to community answers one question but leaves others open. The comparison communities have themselves been shaped by a public policy that answered many of the social problems of racism and poverty—untreated mental illness, enduring joblessness, school failure—with incarceration (Garland 2020; Beckett and Western 2001). Failures of health, employment, or education policy in poor communities have been seen as rooted in a policy outlook that was punitive and viewed residents of poor communities as undeserving (Soss, Fording, and Schram 2011; Katz 2013). From this perspective, prison and community have both been shaped by a policy orientation that is suspicious of the moral worth of poor people, and poor people of color in particular. The short distance between prison and community along the dimensions of victimization, mortality, untreated mental illness, and infectious disease may not reflect the beneficence of prison, but the malignance of a public policy that takes a punitive approach both to crime and alleviating poverty.

Finally, prison conditions—participation in rehabilitation programs, institutional violence, infectious disease, and extreme isolation—can be understood as measures of what Liebling (2004) calls "moral performance" that indicate the dignity of incarcerated people and decent treatment by authorities. Researchers may hesitate to weigh the moral status of prisons in their analysis, leaving that job to policymakers and philosophers. But setting aside the moral question does not eliminate it; any evidence-based recommendation on penal policy also includes a moral stance on the prison. The warehousing, institutional violence, disease, and isolation that are common in American prisons are experienced overwhelmingly by Black and Brown men from low-income communities. In this context, mass incarceration doesn't just influence crime and life chances, but forms part of a moral landscape in which the struggle for dignity follows the contours of poverty and racial inequality.

■ *Thanks to Daniel Nagin, Jessica Simes, Jonathan Simon, Emily Wang, and the editors who provided comments on the paper. Lauren Brinkley-Rubinstein and Kathryn Nowotny generously provided data from the COVID Prison Project. This research was supported by National Science Foundation Grant SES-1823846/1823854.*

References

Aizer, Anna, and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130 (2): 759–803.

American Friends Service Committee. 1971. *Struggle of Justice: A Report on Crime and Punishment in America.* New York: Hill and Wang.

American Psychiatric Association. 2012. "Position Statement on Segregation of Prisoners with Mental Illness." http://nrcat.org/storage/documents/apa-statement-on-segregation-of-prisoners-with-mental-illness.pdf.

Andrews, Don A., James Bonta, and Robert D. Hoge. 1990. "Classification for Effective Rehabilitation: Rediscovering Psychology." *Criminal Justice and Behavior* 17 (1): 19–52.

Apel, Robert, and Gary Sweeten. 2010. "The Impact of Incarceration on Employment during the Transition to Adulthood." *Social Problems* 57 (3): 448–79.

Arrigo, Bruce A., and Jennifer Leslie Bullock. 2008. "The Psychological Effects of Solitary Confinement on Prisoners in Supermax Units: Reviewing What We Know and Recommending What Should Change." *International Journal of Offender Therapy and Comparative Criminology* 52 (6): 622–40.

Beaudry, Gabrielle, Shaoling Zhong, Daniel Whiting, Babak Javid, John Frater, and Seena Fazel. 2020. "Managing Outbreaks of Highly Contagious Diseases in Prisons: A Systematic Review." *BMJ Global Health* 5 (11).

Beck, Allen J. 2015. *Use of Restrictive Housing in US Prisons and Jails, 2011–12.* Washington DC: US Department of Justice.

Beckett, Katherine, and Bruce Western. 2001. "Governing Social Marginality: Welfare, Incarceration, and the Transformation of State Policy." In *Mass Imprisonment: Social Causes and Consequences*, edited by David Garland, 35–50. London: Sage Publications.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2020. "Incarceration, Recidivism, and Employment." *Journal of Political Economy* 128 (4): 1269–1324.

Bick, Joseph A. 2007. "Infection Control in Jails and Prisons." *Clinical Infectious Diseases* 45 (8): 1047–55.

Binswanger, Ingrid A., Patrick J. Blatchford, Shane R. Mueller, and Marc F. Stern. 2013. "Mortality after Prison Release: Opioid Overdose and Other Causes of Death, Risk Factors, and Time Trends from 1999 to 2009." *Annals of Internal Medicine* 159: 592–600.

Binswanger, Ingrid A., Patrick M. Krueger, and John F. Steiner. 2009. "Prevalence of Chronic Medical Conditions among Jail and Prison Inmates in the USA Compared with the General Population." *Journal of Epidemiology and Community Health* 63 (11): 912–19.

Bitney, Kristofer, Elizabeth Drake, Joshua Grice, Michael Hirsch, and Stephanie Lee. 2017. "The Effectiveness of Reentry Programs for Incarcerated Person: Findings for the Washington Statewide Reentry Council." https://www.researchgate.net/publication/329281817_The_Effectiveness_of_Reentry_Programs_for_Incarcerated_Persons_Findings_for_the_Washington_Statewide_Reentry_Council.

Bottoms, Anthony E. 1999. "Interpersonal Violence and Social Order in Prisons." *Crime and Justice* 26: 205–81.

Bowker, Lee H. 1980. *Prison Victimization.* New York: Elsevier.

Bozick, Robert, Jennifer Steele, Lois Davis, and Susan Turner. 2018. "Does Providing Inmates with Education Improve Postrelease Outcomes? A Meta-Analysis of Correctional Education Programs in the United States." *Journal of Experimental Criminology* 14: 389–428.

Brinkley-Rubinstein, Lauren, Josie Sivaraman, David L. Rosen, David H. Cloud, Gary Junker, Scott Proescholdbell, Meghan E. Shanahan, and Shabbar I. Ranapurwala. 2019. "Association of Restrictive Housing during Incarceration with Mortality after Release." *JAMA Network Open* 2 (10).

Bureau of Justice Statistics. 1994. Survey of Inmates of State Correctional Facilities, 1986. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR8711. https://doi.org/10.3886/ICPSR08711.v2.

Bureau of Justice Statistics. 2004a. Survey of Inmates of State Correctional Facilities, 1991. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR6068. https://doi.org/10.3886/ICPSR06068.v1.

Bureau of Justice Statistics. 2004b. Survey of Inmates of State and Federal Correctional Facilities, 1997. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR2598. https://doi.org/10.3886/ICPSR02598.v1.

Bureau of Justice Statistics. 2006. Survey of Inmates of State and Federal Correctional Facilities, 1997.

Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR2598. https://doi.org/10.3886/ICPSR02598.v1.

Bureau of Justice Statistics. 2019. Survey of Inmates of State and Federal Correctional Facilities, 2004. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR4572. https://doi.org/10.3886/ICPSR04572.v6.

Bureau of Justice Statistics. 2020. National Crime Victimization Survey, 2016. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR37296. https://doi.org/10.3886/ICPSR37296.v2.

Bureau of Justice Statistics. 2021. Survey of Prison Inmates, United States, 2016. Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR4572. https://doi.org/10.3886/ICPSR37692.v4.

Bushway, Shawn D., and Raymond Paternoster. 2009. "The Impact of Prisons on Crime." In *Do Prisons Make Us Safer? The Benefits and Costs of the Prison Boom*, edited by Steven Raphael and Michael A. Stoll, 119–50. New York: Russell Sage Foundation.

Carson, E. Ann, and Mary P. Cowhig. 2020. *Mortality in State and Federal Prisons, 2001–2016 – Statistical Tables*. Washington, DC: US Department of Justice.

Chen, M. Keith, and Jesse M. Shapiro. 2007. "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach." *American Law and Economics Review* 9 (1): 1–29.

Cloud, David H., Ernest Drucker, Angela Browne, and Jim Parsons. 2015. "Public Health and Solitary Confinement in the United States." *American Journal of Public Health* 105 (1): 18–26.

Cook, Philip J., Songman Kang, Anthony A. Braga, Jens Ludwig, and Mallory E. O'Brien. 2015. "An Experimental Evaluation of a Comprehensive Employment-Oriented Prisoner Re-Entry Program." *Journal of Quantitative Criminology* 31: 355–82.

Cooley, Dennis. 1993. "Criminal Victimization in Male Federal Prisons." *Canadian Journal of Criminology* 35 (4): 479–95.

Cullen, Francis T. 2004. "The Twelve People Who Saved Rehabilitation: How the Science of Criminology Made a Difference: The American Society of Criminology 2004 Presidential Address." *Criminology* 43 (1): 1–42.

Cullen, Francis T. 2013. "Rehabilitation: Beyond Nothing Works." *Crime and Justice* 42: 299–376.

Cullen, Francis T. 2017. "Correctional Rehabilitation." *Academy for Justice* 4: 235–60.

Delaney, Ruth, Ram Subramanian, Alison Shames, and Nicholas Turner. 2018. *Reimagining Prison*. New York: Vera Institute of Justice.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2019. "The Intergenerational Effect of Parental Incarceration." NBER Working Paper 24186.

Donohue, John J. 2009. "Assessing the Relative Benefits of Incarceration: Overall Changes and the Benefits on the Margin." In *Do Prisons Make Us Safer?: The Benefits and Costs of the Prison Boom*, edited by Steven Raphael and Michael A. Stoll, 269–342. New York: Russell Sage Foundation.

Durlauf, Steven N., and Daniel S. Nagin. 2011. "The Deterrent Effect of Imprisonment." In *Controlling Crime: Strategies and Tradeoffs*, edited by Philip J. Cook, Jens Ludwig, and Justin McCrary, 43–94. Chicago: University of Chicago Press.

Fazel, Seena, and Jacques Baillargeon. 2011. "The Health of Prisoners." *Lancet* 377 (9769): 956–65.

Flamm, Michael W. 2005. *Law and Order: Street Crime, Civil Unrest, and the Crisis of Liberalism in the 1960s*. New York: Columbia University Press.

Gaes, Gerald G., Timothy J. Flanagan, Lawrence L. Motiuk, and Lynn Stewart. 2000. "Adult Correctional Treatment." *Crime and Justice* 26: 361–426.

Garland, David. 2001. *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago: University of Chicago Press.

Garland, David. 2020. "Penal Controls and Social Controls: Toward a Theory of American Penal Exceptionalism." *Punishment & Society* 22 (3): 321–52.

Goffman, Erving. 1961. *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. New York, NY: Anchor Books.

Gough, Ethan, Mirjam C. Kempf, Laura Graham, Marvin Manzanero, Edward W. Hook, Al Bartolucci, and Eric Chamot. 2010. "HIV and Hepatitis B and C Incidence Rates in US Correctional Populations and High Risk Groups: A Systematic Review and Meta-Analysis." *BMC Public Health* 10: 1–14.

Grassian, Stuart. 1983. "Psychopathological Effects of Solitary Confinement." *American Journal of Psychiatry* 140 (11): 1450–54.

Grogger, Jeffrey. 1995. "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics* 110 (1): 51–71.

Guetzkow, Joshua, and Eric Schoon. 2015. "If You Build It, They Will Fill it: The Consequences of Prison Overcrowding Litigation." *Law & Society Review* 49 (2): 401–32.

Haney, Craig. 2006. *Reforming Punishment: Psychological Limits to the Pains of Imprisonment.* Washington, DC: American Psychological Association.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. 2018. "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment." *American Journal of Sociology* 124: 49–110.

Helland, Eric, and Alexander Tabarrok. 2007. "Does Three Strikes Deter? A Nonparametric Estimation." *Journal of Human Resources* 42: 309–30.

Hjalmarsson, Randi and Mathew J. Lindquist. 2020. "The Health Effects of Prison." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3688168.

Hochstetler, Andy, Daniel S. Murphy, and Ronald L. Simons. 2004. "Damaged Goods: Exploring Predictors of Distress in Prison Inmates." *Crime & Delinquency* 50: 436–57.

Huh, Kil, Alex Boucher, Frances McGaffey, Matt McKillop, and Maria Schiff. 2017. *Prison Health Care: Costs and Quality.* Washington DC: Pew Charitable Trusts.

Jung, Haeil. 2014. "Do Prison Work-Release Programs Improve Subsequent Labor Market Outcomes? Evidence from the Adult Transition Centers in Illinois." *Journal of Offender Rehabilitation* 53 (5): 384–402.

Kapoor, Reena, and Robert Trestman. 2016. "Mental Health Effects of Restrictive Housing." In *Restrictive Housing in the US: Issues, Challenges, and Future Directions*, edited by National Institute of Justice, 199–232. Washington DC: National Institute of Justice.

Katz, Michael B. 2013. *The Undeserving Poor: America's Enduring Confrontation with Poverty.* 2nd edition. Oxford: Oxford University Press.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76.

Levitt, Steven D. 1996. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation." *Quarterly Journal of Economics* 111 (2): 319–51.

Liebling, Alison. 1999. "Prison Suicide and Prisoner Coping." *Crime and Justice* 26: 283–359.

Liebling, Alison. 2004. *Prisons and Their Moral Performance: A Study of Values, Quality, and Prison Life.* Oxford: Oxford University Press.

Liman Program and ASCA. 2015. Time-In-Cell: The ASCA-Liman 2014 National Survey of Administrative Segregation in Prison. New Haven, CT: Yale Law School.

Lipton, Douglas, Robert Martinson, and Judith Wilks. 1975. *The Effectiveness of Correctional Treatment: A Survey of Treatment Evaluation Studies.* New York: Praeger.

Lofstrom, Magnus, and Steven Raphael. 2016. "Incarceration and Crime: Evidence from California's Public Safety Realignment Reform." *ANNALS of the American Academy of Political and Social Science* 664 (1): 196–220.

Lopoo, Leonard M., and Bruce Western. 2005. "Incarceration and the Formation and Stability of Marital Unions." *Journal of Marriage and the Family* 67 (3): 721–34.

Lynch, Mona. 2009. *Sunbelt Justice: Arizona and the Transformation of American Punishment.* Stanford: Stanford University Press.

MacKenzie, Doris Layton. 2006. *What Works in Corrections: Reducing the Criminal Activities of Offenders and Delinquents.* Cambridge: Cambridge University Press.

Mariner, Joanne. 2001. *No Escape: Male Rape in U.S. Prisons.* New York: Human Rights Watch.

Martin, Susan E., Lee Sechrest, and Robin Redner, eds. 1981. *New Directions in the Rehabilitation of Criminal Offenders.* Washington DC: National Academy Press.

Mears, Daniel P., and William D. Bales. 2009. "Supermax Incarceration and Recidivism." *Criminology* 47 (4): 1131–66.

Metzner, Jeffrey L., and Jamie Fellner. 2013. "Solitary Confinement and Mental Illness in US Prisons: A Challenge for Medical Ethics." In *Health and Human Rights in a Changing World*, edited by Michael Grodin, Daniel Tarantola, George Annas, and Sofia Gruskin, vol. 1. New York: Routledge.

Morgan, Rachel E., and Jennifer L. Truman. 2020. "Criminal Victimization, 2019." Bureau of Justice Statistics NCJ 255113.

Muller, Christopher. 2018. "Freedom and Convict Leasing in the Postbellum South." *American Journal of Sociology* 124 (2): 367–405.

Mumola, Christopher J. 2005. "Suicide and Homicide in State Prisons and Local Jails." Bureau of Justice Statistics NCJ 210036.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson. 2009. "Imprisonment and Reoffending." *Crime and Justice* 38: 115–200.

National Commission on Correctional Health Care. 2016. "Position Statement: Solitary Confinement (Isolation)." https://www.ncchc.org/solitary-confinement.

Neal, Derek, and Armin Rick. 2014. "The Prison Boom and the Lack of Black Progress after Smith and Welch." NBER Working Paper 20283.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2021. "The Effects of Parental and Sibling Incarceration: Evidence from Ohio." *American Economic Review* 111 (9): 2926–63.

Office of Justice Programs. 2021. "National Institute of Justice: CrimeSolutions." https://crimesolutions.ojp.gov. (accessed February 21, 2021.

Oshinsky, David M. 1997. *Worse Than Slavery: Parchman Farm and the Ordeal of Jim Crow Justice*. New York: Simon and Schuster.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937–75.

Patterson, Evelyn J. 2010. "Incarcerating Death: Mortality in U.S. State Correctional Facilities, 1985–1998." *Demography* 47: 587–607.

Perkinson, Robert. 2010. *Texas Tough: The Rise of America's Prison Empire*. New York: Metropolitan Books.

Phelps, Michelle S. 2011. "Rehabilitation in the Punitive Era: The Gap between Rhetoric and Reality in U.S. Prison Programs." *Law and Society Review* 45 (1): 33–68.

Polanyi, Karl. 1957. *The Great Transformation: The Political and Economic Origins of Our Time*. Boston: Beacon Press.

Puglisi, Lisa B., Giovanni S.P. Malloy, Tyler D. Harvey, Margaret L. Brandeau, and Emily A. Wang. 2021."Estimation of Covid-19 Basic Reproduction Ratio in a Large Urban Jail in the United States." *Annals of Epidemiology* 53: 103–105.

Raphael, Steven, and Michael A. Stoll. 2013a. *Why Are So Many Americans in Prison?* New York: Russell Sage Foundation.

Raphael, Steven, and Michael A. Stoll. 2013b. "Assessing the Contribution of the Deinstitutionalization of the Mentally Ill to Growth in the Us Incarceration Rate." *Journal of Legal Studies* 42 (1): 187–222.

Rantala, Ramona R. 2018. "Sexual Victimization Reported by Adult Correctional Authorities, 2012–15." Bureau of Justice Statistics NCJ 251146.

Reiter, Keramet. 2016. *23/7: Pelican Bay Prison and the Rise of Long-Term Solitary Confinement*. New Haven, CT: Yale University Press.

Rosen, David L, David A. Wohl, and Victor J Schoenbach. 2011. "All-Cause and Cause-Specific Mortality among Black and White North Carolina State Prisoners, 1995–2005." *Annals of Epidemiology* 21 (10): 719–26.

Rothman, David J. 2002a. *The Discovery of the Asylum: Social Order and Disorder in the New Republic*. London: Routledge.

Rothman, David J. 2002b. *Conscience and Convenience: The Asylum and Its Alternatives in Progressive America*. London: Routledge.

Sampson, Robert J., and John H. Laub. 1990. "Crime and Deviance over the Life Course: The Salience of Adult Social Bonds." *American Sociological Review* 55 (5): 609–627.

Sampson, Robert J., and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

Schiff, Maria, Samantha Chao, Kil Huh, Ellyon Bell, Kavita Choudhry, and Matt McKillop. 2014. *State Prison Health Care Spending: An Examination*. Washington DC: Pew Charitable Trusts.

Schnittker, Jason, Michael Massoglia, and Christopher Uggen. 2012. "Out and Down: Incarceration and Psychiatric Disorders." *Journal of Health and Social Behavior* 53 (4): 448–64.

Simon, Jonathan. 2014. *Mass Incarceration on Trial: A Remarkable Court Decision and the Future of Prisons in America*. New York: New Press.

Soss, Joe, Richard C. Fording, and Sanford F. Schram. 2011. *Disciplining the Poor: Neoliberal Paternalism and the Persistent Power of Race*. Chicago: University of Chicago Press.

Struckman-Johnson, Cindy, and David Struckman-Johnson. 2000. "Sexual Coercion Rates in Seven Midwestern Prison Facilities for Men." *The Prison Journal* 80 (4): 379–90.

Subramanian, Ram, and Alison Shames. 2013. *Sentencing and Prison Practices in Germany and the*

Netherlands: Implications for the United States. New York: Vera Institute of Justice.

Sundt, Jody, Emily J. Salisbury, and Mark G. Harmon. 2016. "Is Downsizing Prisons Dangerous? The Effect of California's Realignment Act on Public Safety." *Criminology & Public Policy* 15 (2): 315–41.

Sykes, Gresham M. 1958. *The Society of Captives: A Study of a Maximum Security Prison*. Princeton: Princeton University Press.

Toch, Hans. 1977. *Living in Prison: The Ecology of Survival*. New York: Free Press.

Travis, Jeremy, Bruce Western, and Stephens Redburn, eds. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences*. Washington, DC: National Academy Press.

Turney, Kristin, and Rebecca Goodsell. 2018. "Parental Incarceration and Children's Wellbeing." *Future of Children* 28 (1): 147–64.

Uggen, Christopher. 2000. "Work as a Turning Point in the Life Course of Criminals: A Duration Model of Age, Employment, and Recidivism." *American Sociological Review* 65 (4): 529–46.

United Nations General Assembly. 2015. *United Nations Standard Minimum Rules for the Treatment of Prisoners*. https://www.unodc.org/documents/justice-and-prison-reform/Nelson_Mandela_Rules-E-ebook.pdf.

Wakefield, Sara, and Christopher Wildeman. 2013. *Children of the Prison Boom: Mass Incarceration and the Future of American Inequality*. Oxford: Oxford University Press.

Walmsley, Roy. 2019. "World Prison Population List, Twelfth Edition." https://www.prisonstudies.org/sites/default/files/resources/downloads/wppl_12.pdf.

Wang, Emily A., Bruce Western, Emily P. Backes, and Julie Schuck, eds. 2020. *Decarcerating Correctional Facilities during COVID-19*. Washington DC: National Academies Press.

Weaver, Vesla M. 2007. "Frontlash: Race and the Development of Punitive Crime Policy." *Studies in American Political Development* 21 (2): 230–65.

Weber, Max. 1978. *Economy and Society*, vol. I. Berkeley: University of California Press.

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

Wildeman, Christopher, and Lars H. Andersen. 2020. "Solitary Confinement Placement and Post-Release Mortality Risk among Formerly Incarcerated Individuals: A Population-Based Study." *Lancet Public Health* 5 (2): 107–113.

Wildeman, Christopher, Margaret E. Noonan, Daniela Golinelli, E. Ann Carson, and Natalia Emanuel. 2016. "State-Level Variation in the Imprisonment-Mortality Relationship, 2001–2010." *Demographic Research* 34: 359–72.

Wolff, Nancy, Cynthia L. Blitz, Jing Shi, Ronet Bachman, and Jane A. Siegel. 2006. "Sexual Violence inside Prisons: Rates of Victimization." *Journal of Urban Health* 83 (5): 835–48.

Wooldredge, John D. 1999. "Inmate Experiences and Psychological Well-Being." *Criminal Justice and Behavior* 26 (2): 235–50.

Zweig, Janine M., Jennifer Yahner, Christy A. Visher, and Pamela K. Lattimore. 2015. "Using General Strain Theory to Explore the Effects of Prison Victimization Experiences on Later Offending and Substance Use." *The Prison Journal* 95 (1): 84–113.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 123–146

Rising Geographic Disparities in US Mortality

Benjamin K. Couillard, Christopher L. Foote, Kavish Gandhi, Ellen Meara, and Jonathan Skinner

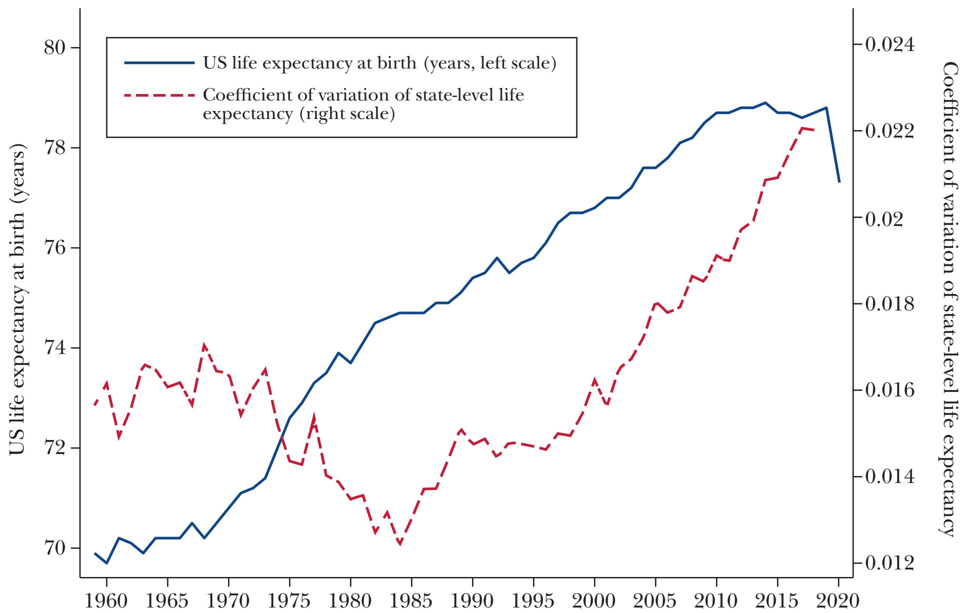
Until recently, Americans could expect to live longer than their parents. Overall US life expectancy rose steadily from the 1960s through the early 2000s. As Figure 1 shows, the 1.5-year drop in life expectancy in 2020 signaled a sharp reversal; indeed, it was the largest decline since World War II. But even before the Covid-19 pandemic, US life expectancy was essentially flat for about a decade and had even declined slightly after 2014. Public health officials and health researchers have become increasingly concerned about this plateau, and, as they studied it, another important fact has emerged: disparities in mortality have become increasingly apparent among different groups in the population.

Much of the recent research on life expectancy focuses on particularly worrisome mortality trends for persons at midlife, defined as ages 25–64. A recent report from the National Academy of Sciences, Engineering, and Medicine (2021) reviews this work and links high and rising midlife mortality rates to two main factors. First, rapid progress that had been made in reducing mortality from some major causes, most notably heart disease, stalled after 2010. Second, deaths from suicide, drug

■ *Benjamin K. Couillard is a PhD student in Economics, University of Toronto, Toronto, Canada. Christopher L. Foote is a Senior Economist and Policy Adviser, Federal Reserve Bank of Boston, Boston, Massachusetts. Kavish Gandhi is a Research Assistant, Federal Reserve Bank of Boston, Boston, Massachusetts. Ellen Meara is Professor of Health Economics and Policy, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. Jonathan Skinner is a Research Professor in Economics, Dartmouth College, Hanover, New Hampshire. Meara and Skinner are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are bcouillard22@gmail.com, chris.foote@bos.frb.org, kavish.gandhi@bos.frb.org, emeara@hsph.harvard.edu, and jon.skinner@dartmouth.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.123.

Figure 1
US Life Expectancy at Birth and State-Level Dispersion in Income



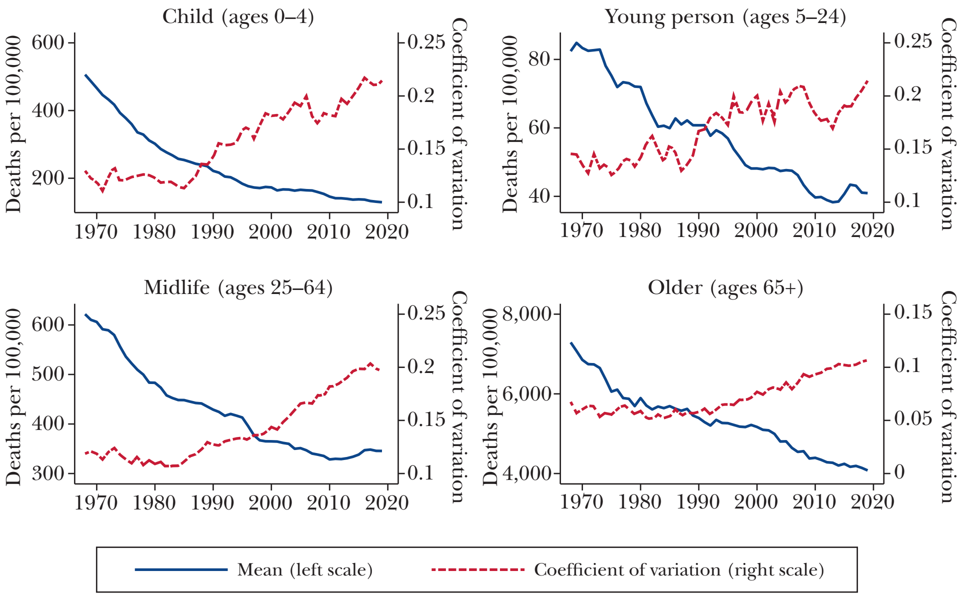
Source: The national life-expectancy rate (1959–2020) is from the National Center for Health Statistics, and state-level rates (1959–2018) are from the USA Mortality Database (https://usa.mortality.org).
Note: The coefficient of variation is the population-weighted standard deviation of state-level life-expectancy rates divided by the national life-expectancy rate.

poisoning, and alcohol-induced causes have risen sharply. These deaths, often labeled "deaths of despair," have been the focus of extensive research by Anne Case and Angus Deaton (2015, 2017, 2020). Regarding mortality disparities, the NAS report noted large and widening mortality differences based on race, ethnicity, economic status, and geography. For example, recent increases in mortality among Black and Hispanic persons have undone years of progress in addressing high mortality rates among these groups (Harris, Woolf, and Gaskin 2021).

In this paper, we document and analyze rising geographic disparities in health, focusing on the state level. Vierboom, Preston, and Hendi (2019) highlight growing local inequality in longevity after 2000; coastal cities gained while rural Appalachia and the South lagged behind. Among US states, Woolf and Schoomaker (2019) document divergence in life expectancy beginning about a decade earlier. Figure 1 shows that the coefficient of variation of state life-expectancy rates (defined as the standard deviation of these rates divided by the mean) began to rise long before average US life expectancy flattened out.

Figure 2 shows that dispersion in state-level life expectancy has been generated by increased dispersion in mortality throughout the age distribution. For the most part, average group-specific mortality rates have trended downward for each of the

Figure 2
Mortality Rates by Age Group (1968–2019)



Source: CDC Wonder database (https://wonder.cdc.gov/).
Note: Data are population-weighted statistics based on state-level age-adjusted mortality rates, defined as deaths per 100,000 persons. The coefficient of variation is the standard deviation of state-level mortality rates divided by the mean.

four age groups depicted (0–4, 5–24, 25–64, and 65+). The stalling of US life expectancy around 2010 resulted from a flattening out of mortality (or outright increases in mortality) for the three age groups younger than 65. But in each of the four groups, dispersion has generally trended higher during the last several decades, especially for the three youngest groups. Although recent trends in race-specific mortality rates contribute to geographic dispersion in mortality, racial patterns alone do not explain why mortality experiences have become more unequal at the state level. Indeed, state-level dispersion has been rising among Black and White non-Hispanic populations separately, while a declining dispersion trend for Hispanics has recently flattened out (as we show in the online Appendix available with this article at the *JEP* website).

What are the most important drivers of mortality divergence across states? One explanation is that geographic disparities are driven by differences in education levels and labor market prospects (Meara, Richards, and Cutler 2008; Case and Deaton 2015, 2017, 2020). In this view, states with relatively large or quickly growing college shares experienced large gains in life expectancy, because recent health gains have been concentrated among Americans with college degrees. As the mortality "penalty" associated with a non-college education grew over time, states with smaller college-educated populations lagged behind.

A second and possibly related explanation is that greater dispersion in state-level mortality rates has been driven by the rising spatial inequality in income. Income is unevenly distributed across the United States, and after converging for most of the 20th century, regions of the country are now growing apart economically (Ganong and Shoag 2017; Gaubert et al. 2021). Chetty et al. (2016) have documented a strong association between income and mortality in the United States. However, much less is known about the influence of longer-term swings over a quarter-century in growth rates of income or about how changing economic circumstances affect common causes of deaths, such as heart disease and cancer.

A third possibility is that the widening divergence in mortality stems from a portmanteau of "place" effects that are independent of state-level income. We think of these effects as capturing both the health behaviors of individuals who live in a place and the evolving features of the region's overall health environment. Much of the prior literature on regional economic conditions and mortality has focused on "deaths of despair," comparing changes in these deaths to economic shocks over relatively short periods of time (Autor, Dorn, and Hanson 2019; Pierce and Schott 2020; Charles, Hurst, and Schwartz 2018; Ruhm 2017; Hollingsworth, Ruhm, and Simon 2017; Ruhm 2019). In contrast to these short-run mechanisms, health disparities across states may arise from long-run changes in state policies or health "investments" that gradually enhance health and longevity (Montez and Berkman 2014; Montez et al. 2019). Examples of long-run health investments include anti-smoking policies, expansions of Medicaid, income support, and norms around health behaviors.

We use data on mortality, income, health behavior, and health-care quality to test these alternative hypotheses for the growth in state-level disparities. Like the recent NAS report and the work of Case and Deaton, we focus on mortality at midlife. We find that national trends in educational attainment and a rising national correlation between education and mortality ultimately explain little of the increasing importance of place in determining mortality. We do not find evidence that states with the most rapid income *growth* experienced the most rapid mortality decline. Instead, states with relatively high income *levels* over the past several decades have experienced the largest improvements in midlife mortality. Although deaths of despair have contributed to the plateau in US life expectancy, even after their recent growth they account for only about one-sixth of all midlife deaths, and we show that midlife disparities are driven largely by other causes of death. Finally, reviewing the growing literature on "place" and health, we argue that the most promising explanation for our findings involve efforts by high-income states to adopt specific health-improving policies and behaviors since at least the early 1990s. Over time, these efforts reduced mortality in high-income states more rapidly than in low-income states, leading to widening spatial disparities in health.

Education and the Rising Dispersion in State-Level Mortality Rates

In a series of important papers and a recent book, Case and Deaton (2015, 2017, 2020, 2021) have documented the striking differences in mortality rates for

Americans with different levels of education. In considering why the spatial dispersion of midlife mortality rose during the past two decades, we first consider the well-known divergence in mortality for people with and without college degrees. Because states differ in their college-educated population shares, the growing national difference between college and non-college mortality rates would by itself generate disparity in state-level mortality, particularly if college-educated persons tended to migrate to states where college attainment was already high.

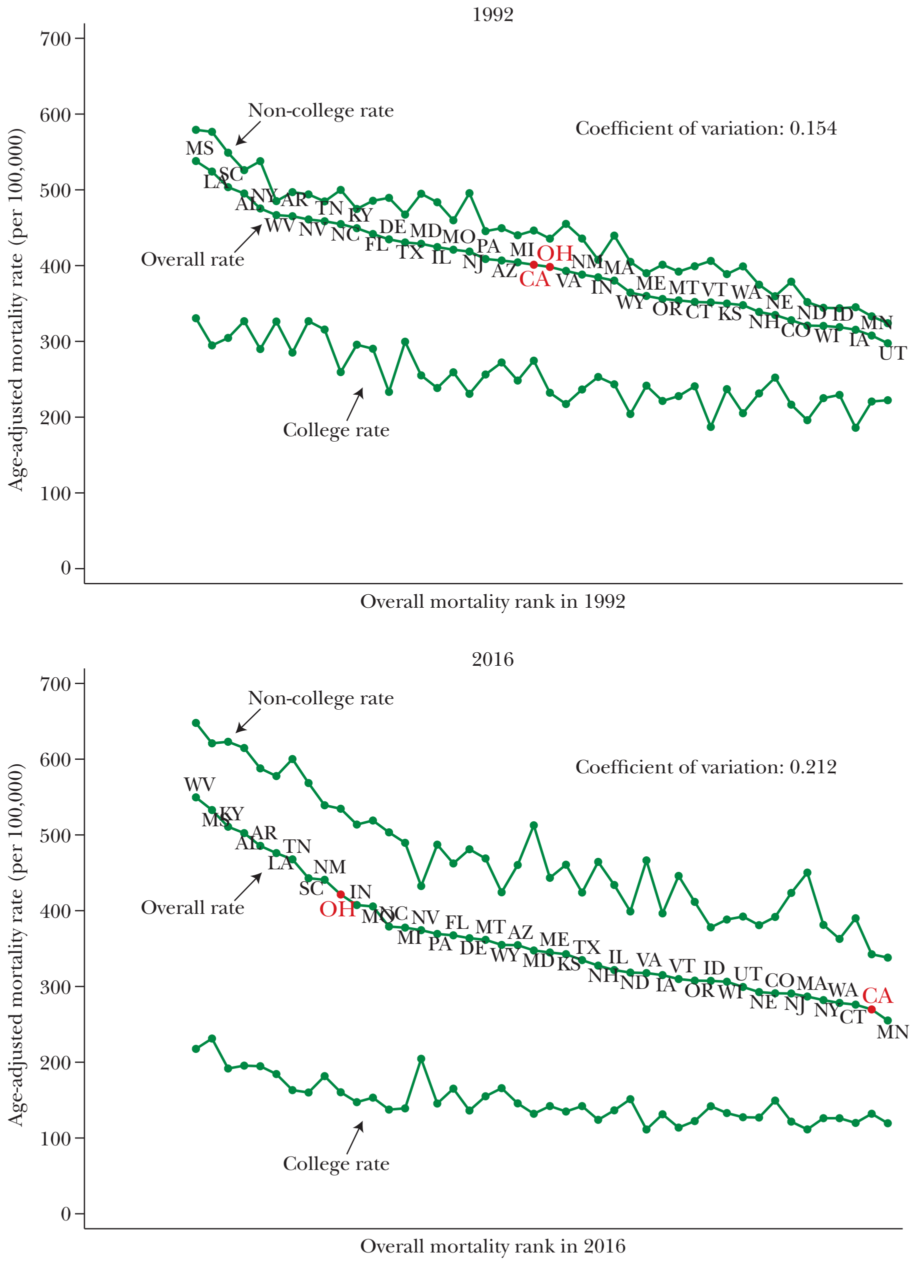
Figure 3 shows all-cause midlife mortality rates separately for 1992 and 2016 for each state, ranked from highest to lowest. The bottom line in each panel is the mortality rate for college-educated residents in each state, while the top line is for non-college; overall state mortality is approximately a weighted average of these two rates, with the weights reflecting the state's share of college-educated residents. Our mortality data come from the collection of individual-level detailed mortality records maintained by the National Center for Health Statistics (NCHS). These records, derived from death certificates, include the cause (or causes) of death for each decedent, as well as demographic information such as age, sex, race, education, and place of residence. Each mortality rate, then, is the number of total deaths divided by the relevant population calculated from the Current Population Survey (CPS) and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER). To account for swings in mortality that would be expected from the aging of large cohorts, like the baby boomers, we age-adjust mortality rates to reflect the deaths that would occur given a fixed age distribution.[1] Starting in 1989, the US Standard Certificate of Death includes a field for the education level of the decedent. Most states were recording education level on death certificates by 1992; the coverage is generally better than 90 percent after 1990 and improves steadily over time.[2] For our cohort of focus, people aged 25 to 64, this is also the population for which the educational information for decedents is most accurate.

As the share of the population with college degrees has grown, overall mortality rates have moved closer to college-educated rates, as shown in Figure 3. The figure also displays the coefficient of variation for overall rates, which has risen from 0.154 to 0.212, an increase of more than one-third. In addition, Figure 3 illustrates the widening gap between college and non-college mortality, a result consistent with Case and Deaton's finding that educational differences in mortality are becoming

[1] We received permission from NCHS to use a restricted-use version of the detailed mortality files, which include state and county of residence, because this field is suppressed in public-use files after 2005. Age-adjustment is done by weighting the raw mortality rates of 10-year age groups in each state and year by shares of population that are constant across states and years. Specifically, the weights used are the standard 2000 reference population weights, drawn from Table V in the Technical Notes of NVSR's "Deaths: Final Data for 2017" (Kochanek et al. 2017).

[2] Four states began collecting education data on death certificates much later: Oklahoma in 1997, Georgia in 2010, South Dakota in 2004, and Rhode Island in 2015; as a result, we omit these states from the analysis. For the remaining states, like other research in this area, the lack of educational information for some decedents requires us to impute this information. Following Case and Deaton (2017), we do this based on the fraction in each education group by year, race, sex, age group, and cause of death; for all-cause mortality, we additionally impute based on state of residence.

Figure 3
Education and Midlife Mortality at the State Level: 1992 and 2016



Source: Authors' calculations using individual-level mortality data from the National Center for Health Statistics.
Note: The middle line in each panel (with accompanying state labels) depicts the all-cause mortality rate for all persons aged 25–64 in the given state and year. The top line depicts the mortality rate for persons in this age group who do not have college degrees, while the lower line depicts the midlife mortality rate of college graduates. Each panel also displays the coefficient of variation for overall mortality in the given year. All mortality rates are age-adjusted. For details, see the online Appendix.

more pronounced over time.[3] For example, in West Virginia and Kentucky in the upper left of the lower panel of Figure 3, all-cause mortality rose between 1992 and 2016 for non-college educated adults, while state-level college-educated mortality rates declined. There is also greater spatial variation in non-college graduate mortality rates, consistent with Chetty et al. (2016) who suggest that spatial variation in mortality is larger for people in lower income groups.

Most importantly, there was considerable movement in state-level mortality rankings between 1992 and 2016. We show a striking example by highlighting California and Ohio in each panel. In 1992, overall mortality rates for these two states were virtually identical. During the 1990s, however, the mortality experiences of the states diverged so that by 2016, the overall mortality rate in California was the second-lowest in the nation, while the rate in Ohio was the 10th highest.

The California–Ohio comparison is consistent with one of the hypotheses discussed earlier: mortality rates in high-education states such as California could have declined by more because of the national mortality trend favoring higher-educated people. We therefore want to ask whether health improved so much in California (and states like it) because these states initially had higher fractions of college-educated adults or because those fractions grew over time.

The role of education in driving mortality dispersion across states can be evaluated with a statistical model. In any given year, a state's overall mortality rate can be thought of as a weighted average of the individual mortality rates for its college-educated and non-college populations, with the weights for this average depending on the state's college-educated population share. In turn, we can think of the state's college mortality rate as the overall mortality trend for all college-educated Americans in that year, plus a state-and-year specific residual. Similarly, the state's non-college rate can be decomposed into the overall national mortality trend for college-educated Americans, plus an additional factor to capture the high (and rising) mortality penalty faced by non-college Americans, plus a non-college state-year residual. For a given year, we can thus characterize each state's mortality rate as:

$$
\begin{aligned}
&\text{Overall state mortality rate (MR)} = (\text{state's college population share})\\
&\quad\times(\text{national college mortality rate} + \text{state's college mortality rate residual})\\
&\quad+(\text{state's non-college population share})\\
&\quad\times(\underbrace{\text{national college mortality rate} + \text{national non-college mortality penalty}}_{\text{national non-college mortality rate}}\\
&\qquad+\text{state's non-college mortality-rate residual}).
\end{aligned}
$$

[3] Our broad measures of college graduates and non-college graduates is likely to mask heterogeneity in educational attainment within these groups: for a discussion of heterogeneity in the non-college group, see Novosad, Rafkin, and Asher (2020).

This framework allows us to allocate the growing state-level divergence in mortality rates across four channels:

(a) *Changes over time in college population shares across states.* These changes could arise from state-level differences in college attendance or from differences in net migration rates of college-educated persons. Because college mortality rates are lower than non-college rates, changes in college shares across states could increase dispersion in overall state mortality.

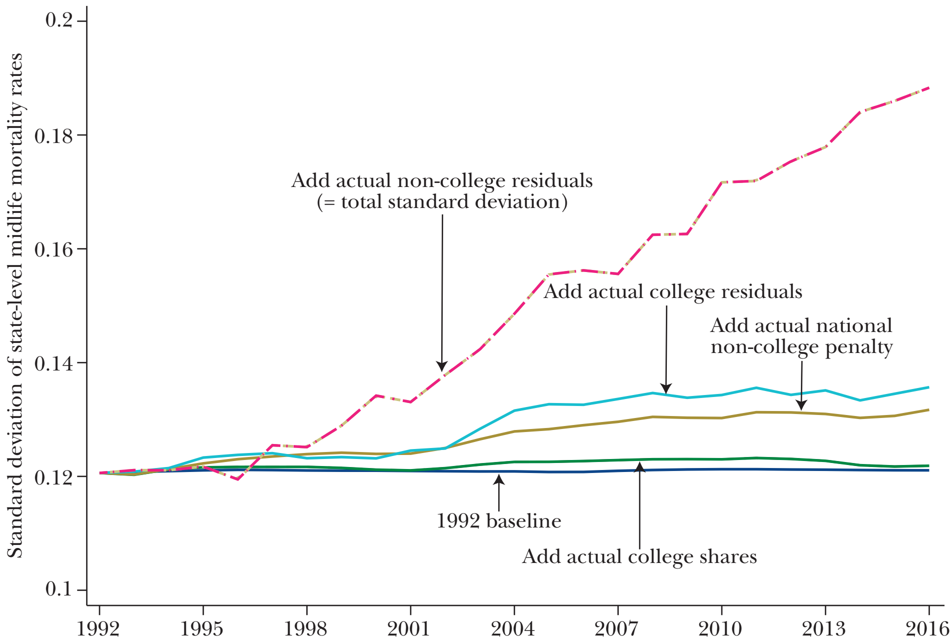
(b) *An increase in the mortality penalty for Americans without a college education.* Holding college shares constant, the well-documented increase in the mortality penalty for non-college Americans would tend to raise relative mortality in states with relatively few college graduates.

(c) *An increase in the standard deviation of the state-level mortality residuals for college residents.* This residual captures any difference between national and state mortality rates among college-educated persons. A gap between a state's college mortality rate and that of the nation could stem from the state's investments in health (interpreted broadly to include public and private health investments), taxes on products that impact health (such as tobacco and alcoholic beverages), and from differential health behaviors. For example, as information about nutrition, exercise, and tobacco's role in health increased, college graduates in states like California may have adopted healthy behaviors more often than college graduates in the nation as a whole.

(d) *An increase in the standard deviation of the state-level mortality residuals for non-college residents.* This term, similar to the college residual, captures the difference between state and national mortality rates among non-college adults. Also, like the college residual, the non-college residual will arise from state-specific policies, taxes, and behaviors that matter for mortality. Examples especially relevant for the non-college population include state-level minimum-wage legislation or the generosity of programs such as Medicaid. State regulations promoting clean air and water could also affect the non-college population disproportionately if these individuals tend to live in environmentally stressed communities.

Figure 4 shows how each of these channels contributes to the growth of state-level dispersion. On the vertical axis is the standard deviation of (log) mortality rates across states, with dispersion rising from 0.12 in 1992 to 0.19 in 2016. The baseline is a flat line because it holds all components of state mortality—college population shares, national mortality rates, and state-and-education specific residuals—constant at their 1992 levels. The other lines in the figure depict standard deviations of the log state mortality rates that are implied when the 1992 values of selected model components are replaced with their actual values. For example, replacing each state's 1992 college-educated population share with its actual evolving college shares (channel a) has only a modest impact on the implied standard deviation of log mortality rates across states, while replacing the 1992 national non-college mortality penalty with the rising actual values of this penalty (channel b) adds a bit more. Combined, however, these two channels account for less than one-sixth of the total increase in state-level standard deviation over time. Rising variation in

Figure 4
Decomposing the Rising Dispersion in State-Level Midlife Mortality Rates (1992–2016)



Source: Authors' calculations using individual-level mortality data from the National Center for Health Statistics.
Note: The dash-dotted line at the top of the figure depicts the actual population-weighted standard deviation of log midlife mortality rates at the state level. The solid line at bottom ("1992 baseline") depicts the constant standard deviation that would have resulted if each of the four components in the model described in the text had remained constant through 2016. The intermediate lines in the figure show the implied standard deviations after progressively adding each of the model's four components in the following order: actual college population shares, the actual national non-college mortality penalty, actual college mortality rate residuals, and actual non-college mortality rate residuals. The panel shows that changes in college shares and the national non-college penalty explain little of the rising standard deviation over time. Most of the increased dispersion is due to widening dispersion in non-college mortality rate residuals. For details, see the online Appendix.

the standard deviation of actual state-level college residuals (channel c) adds an additional 6 percent, but the lion's share is caused by the increase in the standard deviation of residuals for non-college residents (channel d), which accounts for over three-quarters of the overall dispersion.[4]

Why is the contribution of non-college residuals so high? In part, the non-college component is likely to account for a larger share of the standard deviation

[4]This counterfactual varies according to which variables are changed to actual values first, but in the online Appendix we show that our general results are robust to the order in which each channel is introduced.

simply because college-graduates are typically less than 30 percent of the total population.[5] That said, there is independent evidence that variation in mortality among less-educated or lower-income persons is an important reason why mortality rates vary so much geographically. Montez et al. (2019) study education and mortality from the 1980s through 2011, finding that educational differences in mortality across states grew primarily due to divergence among the less-educated groups. Also, Chetty et al. (2016) link mortality records with individual income data to show that across local labor markets, mortality rates vary more at the bottom of the income distribution than at the top.

Yet our model indicates that the geographical variance in non-college mortality rates is not the whole explanation for rising dispersion in state-level mortality. Nor is this rising dispersion a mechanical consequence of the worsening national mortality penalty faced by non-college Americans. Rather, the importance of both residuals in our framework of state-level mortality suggests that in some states, "place effects" have evolved over time to the benefit of both college and non-college residents, and these place effects turn out to be important in explaining why mortality has diverged over the last three decades. An important clue pointing to the importance of place effects is the high within-state correlation of non-college and college residuals produced by the model, which is relatively stable at around 0.70 in both 1992 and 2016. In an extension of this exercise described in the online Appendix, we show that assuming that each state's yearly place effect is an equally weighted average of its non-college and college residuals shows that place effects can explain much of the increased variance attributed to the two sets of residuals in Figure 4.[6] Results such as these suggest that understanding the role of place in health is a key to understanding rising dispersion in health outcomes over time.

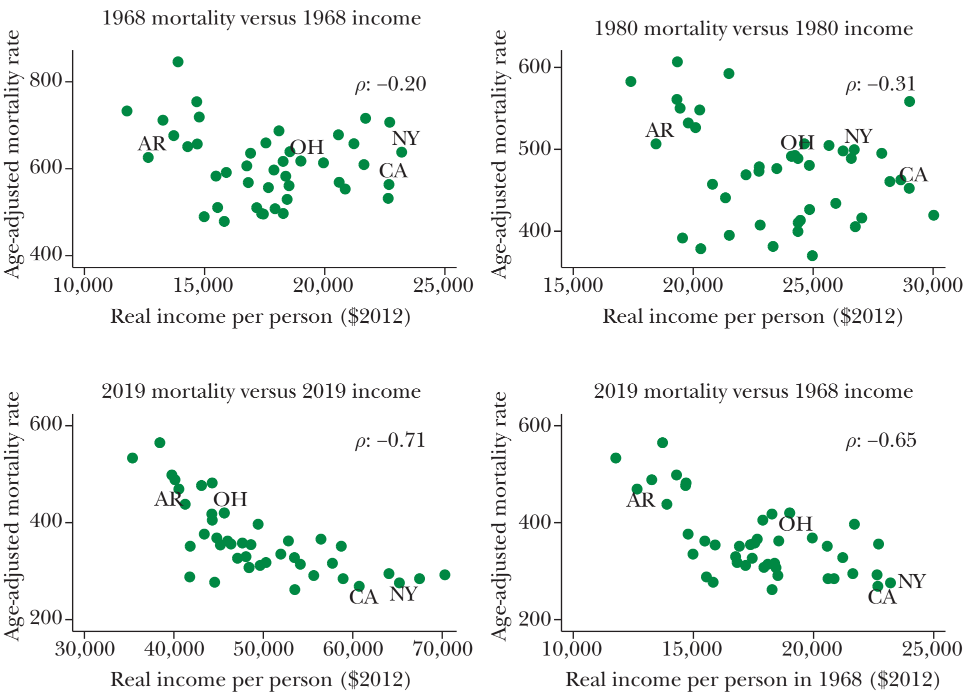
Income and the Rising Divergence of State-Level Mortality Rates

If place effects are large, one may reasonably ask whether other factors associated with mortality are mediated through these effects. An obvious candidate is income, which has been demonstrated at the micro-level to be an important predictor of early mortality (among many others, see Chetty et al. 2016).

In Figure 5, we plot state-specific midlife mortality rates against state-level per capita income in selected years. Our data on income is derived from Census Bureau estimates of total personal income received by the residents of individual states in each year. In this definition, income can include wages and salaries, profits from businesses and farms, payments due to ownership of financial assets, and government transfers, but not capital gains. Per capita income is defined as total personal income divided by state population as of July 1 of

[5] Additionally, because we base our model on the natural log of mortality rates, equal percentage changes in college and non-college rates, combined with higher average rates for non-college populations, will show up as a larger contribution of non-college rates to overall dispersion.
[6] Rather than using an equally weighted average of residuals to create place effects, an alternative method would be to use national shares of college and non-college graduates over the time period considered.

Figure 5
State-Level Income and Midlife Mortality Rates in Selected Years



Source: CDC Wonder database (https://wonder.cdc.gov/).
Note: Data are population-weighted statistics based on state-level age-adjusted mortality rates, defined as deaths per 100,000 persons. The coefficient of variation is the standard deviation of state-level mortality rates divided by the mean.

that year and expressed in 2012 dollars using the price deflator for personal consumption expenditures. Because the mortality rates are not broken down by education, we can rely on public-use mortality data and extend the analysis to 2019.

The upper left panel of the figure plots mortality against income in 1968. In this year, the correlation between mortality and per-capita state income was negligible at –0.20. Residents of New York and California in 1968 had higher average incomes than residents of Arkansas and Ohio (as they do currently), but in that year state-level mortality was similar across all four states. The upper right panel shows that in 1980, the correlation between income and mortality was largely unchanged, even as incomes grew. By 2019, however, state mortality rates had lined up largely in lockstep with income. The lower left panel shows a negative and significant correlation between income and mortality equal to –0.71.

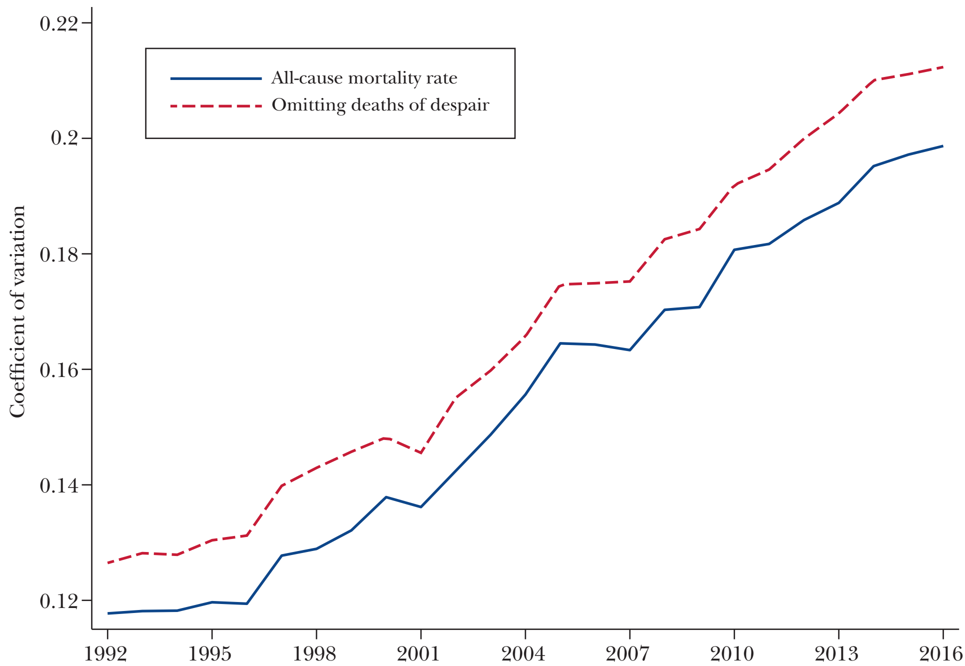
At first glance, a strong correlation between income and mortality in the 2019 cross-section might suggest that changes in economic conditions (like income or unemployment rates) predict changes in mortality. Instead, we find a more subtle pattern. The dramatic lining-up of income and mortality in the lower left panel of Figure 5 was not so much a shift in income rankings across states but rather a reshuffling of state-level place effects. Over time, midlife mortality has become increasingly correlated with the level of income, a result that, except for Pinkovskiy (2019), we had not previously seen. For example, during this period, mortality rates fell rapidly in New York and California while in Ohio and Arkansas they barely budged. Because high-income states in 2019 were typically high-income states in earlier years, we can express the lining-up of mortality and income with income data from previous years, as we do in the lower right panel of the figure. This panel shows that mortality in 2019 is also strongly negatively associated with state-level per capita income from more than 50 years earlier, with a correlation of –0.65. Taken together, these correlations strongly suggest that the greater dispersion of mortality levels across states is not being driven by the growing dispersion of income levels; that is, state-level changes in income do not explain state-level changes in mortality. This result is also supported by other analyses, including Case and Deaton (2017) and Ruhm (2018). Instead, mortality changes have been most favorable in those states that have tended to have high relative levels of income over the past three decades.

An obvious candidate to explain the growing correlation between midlife mortality and income is the growing rate of deaths of despair. Case and Deaton (2017) have not only documented the explosive increase in these deaths during the 21st century but have also shown that spatial dispersion of these deaths has risen dramatically over the same period. As they emphasize, the dramatic growth in midlife mortality is strongly correlated with education. Among college graduates, deaths of despair have remained largely unchanged and show little variation across states. By contrast, deaths of despair in the non-college population have risen sharply, with a particular impact in states such as West Virginia, New Mexico, Ohio, New Hampshire, and Massachusetts (Case and Deaton 2020). If deaths of despair have been concentrated in low-income states, then their recent growth could potentially explain the strengthening correlation between state-level income and mortality that we have documented.

Although deaths of despair have clearly contributed to the widening geographic disparity in mortality rates across states, they are not the primary cause. To see this, note that measured dispersion in midlife mortality has been growing rapidly even when deaths of despair are excluded from the analysis. Figure 6 depicts the coefficient of variation of midlife mortality rates with and without deaths of despair from 1992 to 2016. During this period, the coefficient of variation of mortality rates for deaths excluding deaths of despair increased by 67.9 percent, almost identical to the 68.7 percent increase in variation for all-cause mortality rates.[7]

[7] We acknowledge that deaths of despair are likely understated because of underreporting; a drug overdose might incorrectly be reported as a heart attack (Glei and Preston 2020; Vierboom, Preston, and Hendi 2019). However, the state-level correlation between the growth in deaths of despair, and in other deaths, is just 0.35, so biases are likely to be limited.

Figure 6
State-Level Mortality-Rate Coefficient of Variation with and without Deaths of Despair



Source: Authors' calculations using individual-level mortality data from the National Center for Health Statistics.
Note: Mortality rates are age-adjusted and correspond to persons aged 25-64. Deaths of despair are deaths attributed to cirrhosis (ICD9: 571; ICD10: K70, K73-74), suicide (ICD9: E950-959; ICD10: X60-84, Y87.0), or poisoning (E850-860, E980-982; ICD10: X40-X45, Y10-15).. The coefficients of variation are population-weighted. For details, see the online Appendix.

A key reason that deaths of despair do not completely explain rising dispersion is that even when accounting for their recent rapid growth, these deaths account for only about one-sixth of all deaths at midlife. (Deaths of despair do account for a larger fraction of life-years lost because such deaths tend to occur at younger ages.) The top panel of Figure 7 displays midlife mortality rates in selected years between 1992 to 2016 for deaths of despair and for four of the leading causes of death: cancer (more formally known as "malignant neoplasms"), heart disease, cerebrovascular diseases, and chronic lower respiratory diseases. Not surprisingly, deaths related to cancer and heart disease, the leading causes of death in the United States, are also the most common in 1992. Also notable is the dramatic reduction in death rates for these two diseases, as well as the well-documented (but still unexplained) slowdown in the reduction in heart disease deaths after 2008. Deaths of despair, while more common than cerebrovascular disease and chronic lower respiratory disease, were less common than cancer or heart disease deaths in 1992. By 2016, deaths of despair killed as many Americans aged 25–64 as did heart disease, but fewer than cancer.

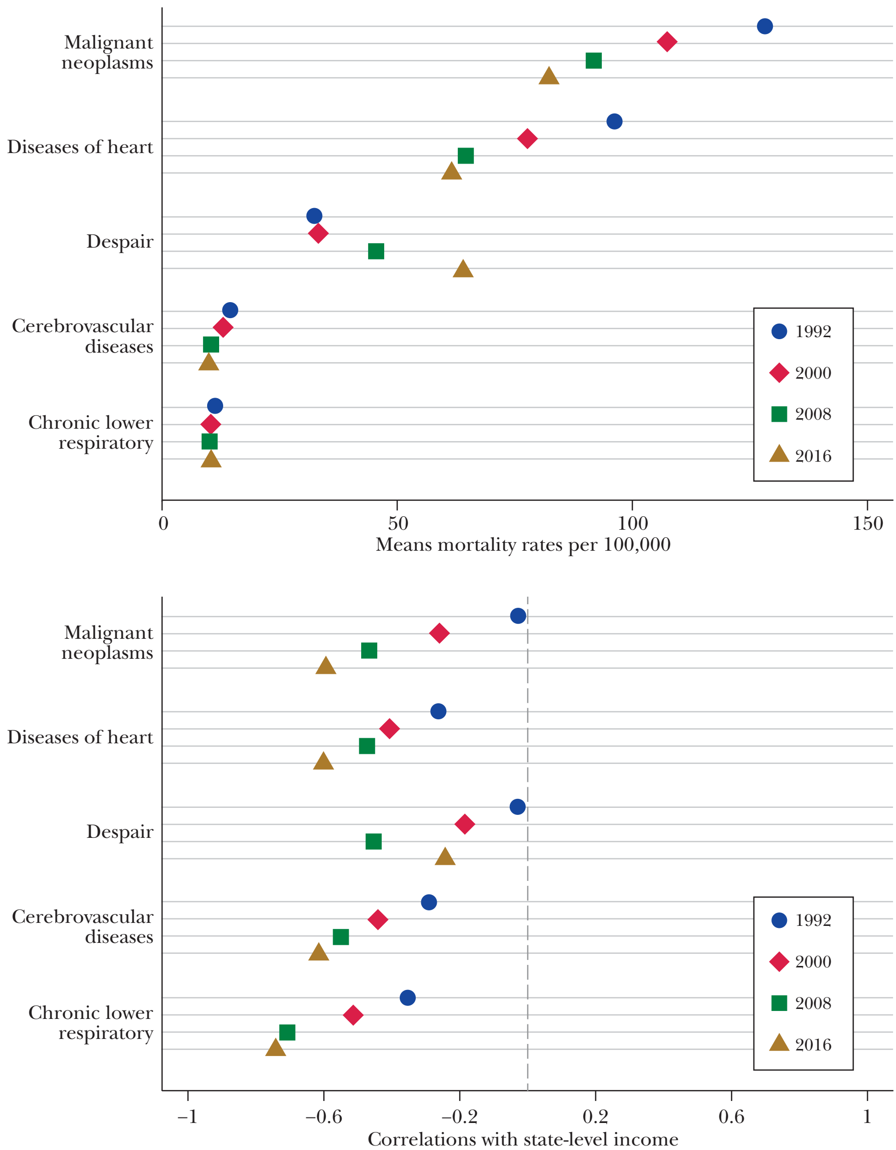
The lower panel of Figure 7 displays the correlation between these causes of death and contemporaneous state income for the same years. Across all causes in the figure, state-level income became more negatively correlated with death rates from 1992 to 2016. Yet while income correlation for deaths of despair follows this pattern for most of the 1990s and early 2000s, the correlation later reverses course and becomes less negative. It is likely that the introduction of fentanyl and other synthetic opioids in recent years have changed the nature of the overdose crisis in the United States, weakening the correlation between state-level income and deaths of despair in the process.

Our earlier example of Ohio and California from Figure 3 helps to illustrate how deaths of despair relate to overall patterns of mortality. As noted earlier, mortality rates in Ohio and California were similar in 1992 (401 deaths per 100,000 in California versus 398 in Ohio). Over time, deaths of despair grew by much more in Ohio so that by 2016, Ohio ranked third-highest among all states in these deaths. But overall mortality in Ohio did not change much over this period, as its large increase in deaths of despair (63 per 100,000) was nearly offset by a decline in other deaths (50 per 100,000). California, on the other hand, experienced a significant decline in overall mortality, to just 270 per 100,000 by 2016. This decline resulted from a small increase in deaths of despair (2 per 100,000) that was swamped by a decline in California's other deaths of 133 per 100,000—almost three times the fall in Ohio. All told, for these two states, deaths of despair accounted for about 40 percent of the widening gap, with the much greater decline in other deaths in California responsible for the remainder.

The ultimate relationship between opioid use, deaths of despair, and regional economic conditions is undoubtedly complex. Several papers have found that exogeneous shifts in manufacturing employment tend to raise adverse opioid-related outcomes (Autor, Dorn, and Hanson 2019; Pierce and Schott 2020; Charles, Hurst, and Schwartz 2018). However, the sizes of the estimated effects are too small to explain much of the massive increase in opioid deaths during the last several years, and some evidence suggests that reductions in manufacturing employment reduce mortality from other causes, such as heart disease (Pierce and Schott 2020). Additionally, regional patterns of deaths of despair are strongly influenced by factors that have little to do with changes in a region's economic conditions. These factors include the growing availability of high-grade heroin at low prices (Quinones 2016) or floods of cheap, illicit, and lethal fentanyl into some communities. Indeed, in the twelve months leading up to May 2020, California experienced an alarming surge in overdose deaths in particular communities (Kurle 2021).

When relating previous research on the economic determinants of health to our results in this section, two things are important to keep in mind. The first is that much of the earlier work relates changes in economic conditions to changes in health outcomes. This approach implicitly assumes a stable relationship between economic conditions and health; if incomes in an area decline, then health also declines due to the constant income-health relationship. But as Figure 5 shows, the income-health relationship itself is changing, as income becomes an increasingly powerful predictor of mortality. A second thing to remember is the long-run

Figure 7
Selected Mortality Rates by Cause and Their Correlations with State-Level Income over Time



Source: Authors' calculations using individual-level mortality data from the National Center for Health Statistics.
Note: Mortality rates are age-adjusted and correspond to persons aged 25–64. Means are population-weighted.Deaths of despair are deaths attributed to cirrhosis (ICD9: 571; ICD10: K70, K73-74), suicide (ICD9: E950-959; ICD10: X60-84, Y87.0), or poisoning (E850-860, E980-982; ICD10: X40-X45, Y10-15). The remaining causes of death are malignant neoplasms (ICD9: 140-208; ICD10: C00-C97), diseases of heart (ICD9: 390-398, 402, 404, 410-429; ICD10: I00-I09, I11, I13, I20-I51), cerebrovascular diseases (ICD9: 430-434, 436-438; ICD10: I60-I69), and chronic lower respiratory diseases (ICD9: 490-494, 496; ICD10: J40-J47).

nature of the disparities that we have analyzed. Health in richer states has been improving more than in poorer states for several decades, so the causes of this divergence likely run deeper than short-term fluctuations in employment or income. Case and Deaton (2017) make a similar point regarding deaths of despair, pointing out that neither state-level measures of income nor changes in income predict the recent rise in these deaths. In their view, deaths of despair are rising not because of short-term economic fluctuations, but rather, because of a long-run devaluation of work performed by persons without college degrees. Their recent book (Case and Deaton 2020) catalogs the devastating effects that this devaluation has had on America's social fabric during the last several decades.

To explain the long-run pattern of mortality differences across states—specifically the strengthening correlation between income and mortality—we also adopt a long-run perspective. In the next section, we contend that the association between state-level income and mortality is probably not a true causal relationship. Instead, the strengthening link between mortality and income reflects differences in regional resources, population behavior, and health-related policies that, over time, have contributed to larger mortality declines in richer states than in poorer ones.

A Portmanteau of State-Level Factors

Our framework for thinking about rising dispersion in state-level mortality has two main components. The first is that health in any point in time is largely determined by decisions made in the past, just as an economy's output of goods and services depends largely on the stock of physical capital built up by past investment. Indeed, health economists often use the concept of "health capital" to capture this phenomenon (Grossman 1972; Case and Deaton 2005). Individuals invest in health capital though behaviors such as regular exercise and maintaining a proper diet. Health capital depreciates over time at a rate that increases with age and in response to factors such as poor health behaviors, stress, and physically demanding occupations (Cutler, Meara, and Stewart 2020). The health-capital framework suggests that various factors have long-lasting effects that "come home to roost" in midlife mortality data many decades on. Given the evidence on the long lag time in health behaviors affecting mortality (Fenelon and Preston 2012), we should expect to observe smoking, obesity, pollution, and stress related to adverse economic conditions several decades ago to be gradually reflected in current midlife mortality (Preston, Vierboom, and Stokes 2018).

A second observation useful for understanding health dispersion is that states differ greatly in their health investment and depreciation rates. A classic example of this phenomenon is due to Victor Fuchs (1974), who observed that Utah exhibited much lower mortality than neighboring Nevada, despite similar levels of income, education, and access to health care. Fuchs argued that this gap could be explained by differing *behavior* in the two states, noting that rates of smoking, drinking, and family instability were much lower in Utah (where the majority of residents are members of the Mormon Church) than in Nevada. State-level differences in health investment and depreciation rates can also be influenced by *policies* related to health. States that instituted high cigarette and liquor taxes or that expanded Medicaid under the Affordable Care Act might expect to see reduced rates of smoking and drinking and improved rates of health investment and depreciation among their residents.

Our hypothesis is that the widening divergence in midlife mortality and the tightening relationship between mortality and income reflect the long-run effects of varying behaviors and policies related to health capital during the last several decades. The data suggest that residents of high-income states have enacted policies and adopted behaviors with long-run payoffs to midlife mortality that are becoming increasingly apparent over time.

One question raised by this hypothesis is why health outcomes are diverging so much now—why hasn't health always been better in high-income states than low-income states? In 1992, high-income states were no more likely to experience lower mortality than low-income states. It was certainly not because high-income people at the time were sicker; individual-level analyses using data from the same period demonstrated a strong negative income gradient in mortality (Pappas et al. 1993), and a similar negative relationship between smoking and income was also apparent. Nor can lagged health effects explain this surprising result; unlike the strong link between 1992 income and current state-level mortality, there is only a weak association between state-level income in 1968 and mortality rates in 1992.

To explain why state differences in mortality have become more aligned with state-level variables like income after about 1990, we instead hypothesize that in the middle of the 20th century, social structures in low-income states provided more safeguards against adverse health outcomes. Perhaps more importantly, during this period there may have been more opportunities for risky behavior in high-income states. Black et al. (2015) show that African-Americans who migrated from the Deep South during the Great Migration experienced higher levels of mortality than those who stayed home, conditional on their initial health statuses. Although migrants may have had higher incomes in the North, "beneficial health benefits due to economic and social improvement were apparently swamped by other forces, such as changes in behavioral patterns that were detrimental to long-term health, including higher propensities to smoke and consume alcohol" (p. 501).[8] By the late 20th century, however, high-income states were more likely to enact health investments that over the next quarter-century resulted in more effective safety nets, more rapid diffusion of effective pharmaceutical treatments, a reduction in smoking, and a consequent decline in all-cause mortality (Montez et al. 2019, 2020; Miller and Wherry 2019; Buxbaum et al. 2020).

The hypothesis that investments related to Medicaid matter for the evolution of mortality has empirical support. Several authors, drawing upon different time periods and settings, show important evidence of plausibly causal reductions in

[8] Further evidence on the importance of state policies comes from Kansas, which imposed prohibition in 1880, not ending it until 1948. Perhaps not coincidentally, in 1959, Kansas was tied in first place for the state with the highest life expectancy.

mortality and morbidity linked to state differences in Medicaid policies. Owing to Medicaid eligibility's link with Aid to Families with Dependent Children, a program dating to 1935 (and commonly referred to as "welfare"), there was substantial cross-state variation in the shares of newborns eligible for Medicaid. Using that variation, Goodman-Bacon (2018) estimates that infant mortality fell for newborn cohorts after Medicaid's implementation in the 1960s and 1970s, and it did so in states with higher rates of eligibility for Medicaid. In the aggregate, nonwhite infant mortality fell by 11 percent in relation to Medicaid's implementation, and it did so for the causes of death amenable to medical intervention at that time (Goodman-Bacon 2018). Later expansions of Medicaid (in the late 1980s and early 1990s) to pregnant women and newborns with slightly higher incomes also coincided with reductions in infant mortality (Currie and Gruber 1996). States that expanded eligibility for Medicaid under the Affordable Care Act saw declines in mortality and morbidity among near-elderly adults (Miller, Johnson, and Wherry 2021).

Even more important for the time period we study, the implementation of Medicaid and its later expansions to pregnant low-income women have been linked to lower morbidity and mortality in the long run (Goodman-Bacon 2021; Miller and Wherry 2019). Again, using state variation in eligibility for Medicaid when first implemented due to its link to state participation in the Aid to Families with Dependent Children, Goodman-Bacon (2021) estimates: "Medicaid added 10 million quality adjusted life-years for cohorts born between 1955 and 1975 and saved the government more than twice its original cost" (p. 2588). This latter point is important since states share up to half the Medicaid program costs, so spending more crowds out other beneficial state spending. Later Medicaid expansions of the 1990s also had lasting effects, with infants whose mothers gained Medicaid coverage in the early 1990s experiencing lower rates of chronic conditions or hospitalizations for diabetes and obesity in adulthood (Miller and Wherry 2019).

Other health programs targeting low-income populations matter for the evolution of long-term health, too. Using variation in the opening of Community Health Centers in the 1960s and 1970s (designed to care for medically under-served populations), Bailey and Goodman-Bacon (2015) showed that age-adjusted mortality rates had declined by an additional 2 percent in counties that opened Community Health Centers compared to those that did not. Further, the mortality decline was driven by deaths to adults over age 50. This pattern we see is also consistent with the hypothesis suggested by Case and Deaton (2017) that cohorts entering the workforce in the 1970s and 1980s experienced a changed economic landscape, one which shifted particularly against people without college degrees.

Another important policy for health is environmental policy, since particulate pollution both sickens and kills, especially among vulnerable residents (Deryugina et al. 2019). A recent paper mapped changes in particulate pollution in the United States from 1980 to 2016, to show that particulate pollution has declined everywhere, though not necessarily equally (Colmer et al. 2020). Returning to our example of diverging mortality rates in Ohio and California, it is interesting that pollution declined by more in Ohio than in California during this time period; West Virginia experienced among the greatest improvement in air quality. Thus, policies to reduce particulate pollution seem unlikely to explain this pattern of diverging mortality across states.

Whereas the empirical work cited so far in this section has investigated formal policies, a growing body of research examines differences in informal health-care practices across geographic areas. One example is the riskiness of prescriptions. Finkelstein, Gentzkow, and Williams (2019b) find that Medicare patients moving from regions with low levels of opioid prescriptions to regions with high levels are more likely to receive risky opioid prescriptions in their new communities. More generally, the question of whether the overall quality of health care has been converging or diverging across geographic areas during the past three decades is unresolved (Skinner and Staiger 2015). As discussed above, exposure to Medicaid improves long-term health outcome for children and adults, but quantifying how they explain the variation in this study is a subject of ongoing research.

All told, there is strong empirical support for the notion that specific health-related policies and behaviors differ across states, and that these differences matter for mortality. But quantifying how much of the total rise in state-level mortality dispersion can be explained by a health-capital model is more ambitious due to the long lags between investments and outcomes and the myriad types of policies and behaviors that might be relevant. It is even more difficult to quantify the separate contributions of policies versus behavior, given the likely feedback between these two "inputs" into the health-capital framework.

Even so, the health-capital model can help us understand some puzzles in the empirical literature. For example, one type of behavior—smoking—typically has a far larger effect on mortality than its direct clinical impact would predict (Cutler et al. 2011). Consistent with a broad health-capital model, Montez et al. (2019) observe that the outsized effect of smoking on health in area-level regressions can be understood by noting that changes in smoking behavior are often correlated with changes in health-related policies, including policies unrelated to smoking. In New York, for example, smoking rates in 1992 were 22.1 percent, about the same as North Dakota (21.9 percent) and only slightly below Mississippi (23.6 percent). By 2016, smoking had fallen to 9.2 percent in New York, compared to significantly smaller decreases in North Dakota (14.0 percent) and Mississippi (16.6 percent). Since the early 1980s, New York has imposed a substantial excise tax on cigarettes, which reached $4.35 per pack in 2016. But as Montez et al. argue, the higher cigarette tax in New York was part of a bundle of initiatives which, to one extent or another, tended to improve public health. For example, New York also participated in Medicaid expansion, implemented its own earned income tax credit, and set a minimum wage above the federal level ($9.00 per hour in 2016). In contrast, Mississippi has a negligible cigarette tax ($0.68 per pack in 2016), opted out of Medicaid expansion, does not offer its own earned income tax credit, and defaulted to the federal minimum wage. In addition, Mississippi has preempted local governments from implementing health-promoting legislation, such as paid sick days, a higher minimum wage, stricter firearm regulations, and nutrition labeling in restaurants.

To explore the plausibility of this explanation, we experimented with regressions with state-level mortality as the dependent variable and various explanatory variables, including smoking and obesity rates. To capture state-level economic factors, we include state-level income, poverty rates, and manufacturing employment shares. We also include rates of prescribing effective or risky drugs, intended to capture health-care quality in 2008–2010 (Munson et al. 2013). Of course, these regression results should not be viewed as causal, and even interpreting the coefficients is tricky given the well-understood risks of using aggregated data to make inferences about individual causal factors.[9] Details of these regressions and the underlying data sources are available in the online Appendix.

Here, we simply note two general patterns that emerge. First, consistent with our earlier results on state-level income and mortality, income has a strong negative correlation with mortality in 2016 but no particular relation in 1992. However, when we include the additional control variables, the later income coefficient becomes much less negative. This reduction suggests that high-income states differ from low-income states along a variety of dimensions relevant for health, which are being captured in some ways by the additional controls.

Second, we find that the importance of smoking in these regressions is rising over time, even after controlling for income.[10] This is consistent with interpreting the state-level smoking rate as a "sentinel measure" of midlife mortality, with lower smoking rates reflecting a variety of public health efforts to encourage more healthy behavior. Indeed, one might view these evolving health-related factors proxied for by smoking as the dynamic equivalent of the static Utah-Nevada comparison by Fuchs (1974), in which behavior is influenced by policies, and vice versa.

Conclusion

We have documented a sharp increase in state-level disparities in midlife mortality, a result consistent with an emerging epidemiological literature (Vierboom, Preston, and Hendi 2019; Montez et al. 2019). This divergence has contributed to a more unequal America; West Virginia's midlife mortality rate is nearly double that in Minnesota. These widening geographic disparities in state-level mortality cannot be attributed to changing spatial patterns in education levels, income inequality, or rising deaths of despair. Instead, rising spatial inequality in midlife mortality results from some states experiencing dramatic overall declines in mortality across educational groups, while other states have experienced at best only modest progress. The first-order question is why high-income states have done so much better.

[9] This is sometimes referred to as the "ecological fallacy." As Gelman (2010) points out, the 15 poorest American states voted Republican in 2004, yet an analysis of individual-level data demonstrates a positive association between income and Republican voting.

[10] State-level smoking data come from the Behavioral Risk Factor Surveillance System (BRFSS) from the Centers for Disease Control and Prevention, an annual set of telephone surveys that collects state-level data on health behaviors. We use the BRFSS' post-stratification weights to construct state-level shares of daily smokers and obesity, where daily smokers are defined as respondents who reported smoking every day and having smoked at least 100 cigarettes throughout their lifetime. We also considered obesity, defined as having a body mass index greater than 30.0, but it was much less predictive of mortality. See the online Appendix for further details.

Our review of the evidence indicates that differential adoption of policies such as tobacco taxes, Medicaid expansions, and income support in high-income but not low-income states, have led to both widening spatial disparities in mortality and to an increasingly close negative association between income and mortality. These policies are distinct from but complementary to health-related behaviors that also differ across states.

We are certainly not the first to observe the importance of place for health, and there is a long-standing literature in geography and social epidemiology on the estimation and interpretation of place effects (McLafferty 2020). In the economics literature, there is a growing interest in estimating causal effects of place that abstract from selection effects that arise when, for example, people in poor health move to low-income neighborhoods lacking access to medical care (Jokela 2014). Studies of people who move can adjust for such selection, particularly when moves are randomized or exogenous (Chyn and Katz 2021). For example, randomized housing vouchers (Kling, Liebman, and Katz 2007; Ludwig et al. 2012) caused families receiving public housing vouchers to leave low poverty neighborhoods, while the destruction of large public housing projects (Chyn 2018) induced moves to lower poverty neighborhoods. Deryugina and Molitor (2020) examined older residents of New Orleans, many of whom moved after Hurricane Katrina in 2005. A notable finding was that average 8-year survival for all Medicare beneficiaries living in New Orleans in 2005 was two percentage points higher than expected in the absence of Katrina, even after accounting for residents who remained in New Orleans, or who died due to direct or indirect effects of the hurricane. In a companion paper in this volume, Deryugina and Molitor consider in more detail the mechanisms by which moving to a new region can affect longevity.

The causal place effects identified in the mover studies are conceptually different than the residual place effects we measure in our study; the short-term impact on health of moving from Mississippi to New York is different from the longer-term effects of growing up in Mississippi versus growing up in New York. For example, Finkelstein, Gentzkow, and Williams (2019a) found that the estimated causal effect of moving to a given region was often different from the underlying health of permanent residents. The cumulative effects of regional policies over the life-cycle—Medicaid coverage at birth, parental income support while a child, tobacco and alcohol taxes during adolescence, and higher-quality medical care during adulthood—are thus likely to exert a larger impact on life expectancy than the short-run impact of moving to a new neighborhood and changing physicians.

Going beyond mover studies to identify the determinants of place effects throughout the life cycle will be challenging. In particular, measuring the relative contributions of policies versus behavior on cross-state differences in health parallels the difficulty of disentangling effects of institutions versus culture on cross-country differences in income and wealth. Two proponents of the importance of institutions in development have observed that "England in the nineteenth century was . . . a very unhealthy place, but the government gradually invested in clean water, in the proper treatment of sewage and effluent, and eventually in an effective health service" (Acemoglu and Robinson 2012, p. 51). The authors interpret these improvements not as the cause of England's rapid economic growth, but instead as a consequence of its economic success. Lessons from this literature on institutions have an encouraging policy implication: Although states with high income have shown the way, states with lower income capacity are not inexorably constrained to rates of midlife mortality that rank among the worst in developed countries.

■ *We are grateful to Anne Case, Angus Deaton, Amy Finkelstein, Matt Gentzkow, Erik Hurst, Heidi Williams, and especially Timothy Taylor for helpful comments and suggestions. Meara and Skinner are grateful for financial support from the National Institute on Aging (AG060104-01). The opinions expressed in this paper are those of the authors and do not indicate concurrence by the Federal Reserve Bank of Boston, the principals of the Board of Governors, of the Federal Reserve System.*

References

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Publishing Group.

Autor, David, David Dorn, and Gordon Hanson. 2019. "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men." *American Economic Review: Insights* 1 (2): 161–78.

Bailey, Martha J., and Andrew Goodman-Bacon. 2015. "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans." *American Economic Review* 105 (3): 1067–1104.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review* 105 (2): 477–503.

Buxbaum, Jason D., Michael E. Chernew, A. Mark Fendrick, and David M. Cutler. 2020. "Contributions of Public Health, Pharmaceuticals, and Other Medical Care to US Life Expectancy Changes, 1990–2015." *Health Affairs* 39 (9): 1546–56.

Case, Anne, and Angus Deaton. 2005. "Broken Down by Work and Sex: How Our Health Declines." In *Analyses in the Economics of Aging*, edited by David A. Wise, 185–212. Chicago: University of Chicago Press.

Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112 (49): 15078–83.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity* 2017 (Spring): 397–476.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism*. Princeton, NJ: Princeton University Press.

Case, Anne, and Angus Deaton. 2021. "Life Expectancy in Adulthood Is Falling for Those without a BA Degree, but as Educational Gaps Have Widened, Racial Gaps Have Narrowed." *Proceedings of the National Academy of Science* 118 (11).

Centers for Disease Control and Prevention (CDC). n.d. "Behavioral Risk Factor Surveillance System (BRFSS)." US Department of Health and Human Services, Centers for Disease Control and Prevention.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2018. "The Transformation of Manufacturing and the Decline in US Employment." In *NBER Macroeconomics Annual 2018*, Vol. 33, edited by

Martin Eichenbaum and Jonathan Parker, 307–372. Chicago: University of Chicago Press.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association between Income and Life Expectancy in the United States, 2001–2014." *JAMA* 315 (16): 1750–66.

Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review* 108 (10): 3028–56.

Chyn, Eric, and Lawrence F. Katz. 2021. "Neighborhoods Matter: Assessing the Evidence for Place Effects." NBER Working Paper 28953.

Colmer, Jonathan, Ian Hardman, Jay Shimshack, and John Voorheis. 2020. "Disparities in PM2.5 Air pollution in the United States." *Science* 369 (6503): 575–78.

Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy*, 104 (6): 1263–96.

Cutler, David M., Fabian Lange, Ellen Meara, Seth Richards-Shubik, and Christopher J. Ruhm. 2011. "Rising Educational Gradients in Mortality: The Role of Behavioral Risk Factors." *Journal of Health Economics* 30 (6): 1174–87.

Cutler, David M., Ellen Meara, and Susan Stewart. 2020. "Socioeconomic Status and the Experience of Pain: An Example from Knees." NBER Working Paper 27974.

Dartmouth Atlas Project. 2013. "The Dartmouth Atlas of Medicare Prescription Drug Use." Dartmouth College Report.

Deryugina, Tatyana, and David Molitor. 2020. "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina." *American Economic Review* 110 (11): 3602– 33.

Deryugina, Tatyana, and David Molitor. 2021. "The Causal Effects of Place on Health and Longevity." University of Illinois at Urbana-Champaign Report.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." *American Economic Review* 109 (12): 4178–4219.

Fenelon, Andrew, and Samuel H. Preston. 2012. "Estimating Smoking-Attributable Mortality in the United States." *Demography* 49 (3): 797–818.

Finkelstein, Amy, Matthew Gentzkow, and Heidi L. Williams. 2019a. "Place-Based Drivers of Mortality: Evidence from Migration." NBER Working Paper 25975.

Finkelstein, Amy, Matthew Gentzkow, and Heidi L. Williams. 2019b. "What Drives Prescription Opioid Abuse? Evidence from Migration." NBER Retirement and Disability Research Center NB19-02.

Fuchs, Victor R. 1974. *Who Shall Live? Health, Economics and Social Choice.* New York: Basic Books.

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the US Declined?" *Journal of Urban Economics* 102: 76–90.

Gaubert, Cecile, Patrick Kline, Dami´an Vergara, and Danny Yagan. 2021. "Trends in US Spatial Inequality: Concentrating Affluence and a Democratization of Poverty." NBER Working Paper 28385.

Gelman, Andrew. 2010. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do--Expanded Edition.* Princeton: Princeton University Press.

Glei, Dana A., and Samuel H. Preston. 2020. "Estimating the Impact of Drug Use on US Mortality, 1999-2016." *PLoS One* 15 (1).

Goodman-Bacon, Andrew. 2018. "Public Insurance and Mortality: Evidence from Medicaid Implementation." *Journal of Political Economy* 126 (1): 216–62.

Goodman-Bacon, Andrew. 2021. "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes." *American Economic Review* 111 (8): 2550–93.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.

Harris, Kathleen Mullan, Steven H. Woolf, and Darrell J. Gaskin. 2021. "High and Rising Working-Age Mortality in the US: A Report from the National Academies of Sciences, Engineering, and Medicine." *JAMA* 325 (20): 2045–46.

Hollingsworth, Alex, Christopher J. Ruhm, and Kosali Simon. 2017. "Macroeconomic Conditions and Opioid Abuse." *Journal of Health Economics* 56: 222–33.

Jokela, Markus. 2014. "Are Neighborhood Health Associations Causal? A 10-Year Prospective Cohort Study with Repeated Measurements." *American Journal of Epidemiology* 180 (8): 776–84.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

Kochanek, Kenneth D., Sherry L. Murphy, Jiaquan Xu, and Elizabeth Arias. 2017. "Deaths: Final Data for 2017." *National Vital Statistics Reports* 68 (9).

Kurle, Sydney. 2021. "The Silent Epidemic: Drug Overdose Deaths Surge in California Pre-Pandemic." *State of Reform*, January 15. https://stateofreform.com/featured/2021/01/the-silent-epidemic-drug-overdose-deaths-surge-in-california-pre-pandemic/.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science* 337 (6101): 1505–10. Data Files.

McLafferty, Sara. 2020. "Place and Quantitative Methods: Critical Directions in Quantitative Approaches to Health and Place." *Health and Place* 6.

Meara, Ellen R., Seth Richards, and David M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981–2000." *Health Affairs* 27 (2): 350–60.

Miller, Sarah, Norman Johnson, and Laura R. Wherry. 2021. "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data." *Quarterly Journal of Economics* 136 (3): 1783–1829.

Miller, Sarah, and Laura R. Wherry. 2019. "The Long-Term Effects of Early Life Medicaid Coverage." *Journal of Human Resources* 54 (3): 785–824.

Montez, Jennifer Karas, Jason Beckfield, Julene Kemp Cooney, Jacob M. Grumbach, Mark D. Hayward, Huseyin Zeyd Koytak, Steven H. Woolf, and Anna Zajacova. 2020. "US State Policies, Politics, and Life Expectancy." *Milbank Quarterly* 98 (3): 668–99.

Montez, Jennifer Karas, and Lisa F. Berkman. 2014. "Trends in the Educational Gradient of Mortality among US Adults Aged 45 to 84 years: Bringing Regional Context into the Explanation." *American Journal of Public Health* 104 (1): 82–90.

Montez, Jennifer Karas, Anna Zajacova, Mark D. Hayward, Steven H. Woolf, Derek Chapman, and Jason Beckfield. 2019. "Educational Disparities in Adult Mortality across US States: How Do They Differ, and Have They Changed since the Mid-1980s?" *Demography* 56 (2): 621–44.

Munson, Jeffrey, Nancy E. Morden, David C. Goodman, Luca F. Valle, and John E. Wennberg. 2013. *The Dartmouth Atlas of Medicare Prescription Drug Use*. Lebanon, NH: The Dartmouth Institute for Health Policy and Clinical Practice.

National Academies of Sciences, Engineering, and Medicine. 2021. *High and Rising Mortality Rates among Working-Age Adults*. Washington, DC: The National Academies Press.

Novosad, Paul, Charlie Rafkin, and Sam Asher. 2020. "Mortality Change among Less Educated Americans." https://paulnovosad.com/pdf/nra-mortality.pdf.

Pappas, Gregory, Susan Queen, Wilbur Hadden, and Gail Fisher. 1993. "The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986." *New England Journal of Medicine* 329: 103–109.

Pierce, Justin R., and Peter K. Schott. 2020. "Trade Liberalization and Mortality: Evidence from US Counties." *American Economic Review: Insights* 2 (1): 47–64.

Pinkovskiy, Maxim. 2019. "Does US Health Inequality Reflect Income Inequality—or Something Else?" *Liberty Street Economics*, October 15. https://libertystreeteconomics.newyorkfed.org/2019/10/does-us-health-inequality-reflect-income-inequalityor-something-else/.

Preston, Samuel H., Yana C. Vierboom, and Andrew Stokes. 2018. "The Role of Obesity in Exceptionally Slow US Mortality Improvement." *Proceedings of the National Academy of Sciences* 115 (5): 957–61.

Quinones, Sam. 2016. *Dreamland: The True Tale of America's Opiate Epidemic*. London: Bloomsbury Press.

Ruhm, Christopher J. 2017. "Geographic Variation in Opioid and Heroin Involved Drug Poisoning Mortality Rates." *American Journal of Preventive Medicine* 53 (6): 745–53.

Ruhm, Christopher J. 2018. "Deaths of Despair or Drug Problems?" NBER Working Paper 24188.

Ruhm, Christopher J. 2019. "Drivers of the Fatal Drug Epidemic." *Journal of Health Economics* 64: 25–42.

Skinner, Jonathan, and Douglas Staiger. 2015. "Technology Diffusion and Productivity Growth in Health Care." *Review of Economics and Statistics* 97 (5): 951–64.

Vierboom, Yana C., Samuel H. Preston, and Arun S. Hendi. 2019. "Rising Geographic Inequality in Mortality in the United States." *SSM Population Health* 9.

Woolf, Steven H., and Heidi Schoomaker. 2019. "Life Expectancy and Mortality Rates in the United States, 1959–2017." *JAMA* 322 (20): 1996–2016.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 147–170

The Causal Effects of Place on Health and Longevity

Tatyana Deryugina and David Molitor

High life expectancy is a hallmark of economic development. Between 1960 and 2018, life expectancy at birth in the United States increased by nine years, from 69.7 to 78.7 years (Bastian et al. 2020). This average masks large and persistent geographic differences in mortality and life expectancy within the United States (Murray et al. 2006; Chetty et al. 2016; Currie and Schwandt 2016; Dwyer-Lindgren et al. 2017; Mokdad et al. 2018; Woolf and Schoomaker 2019). For example, life expectancy is more than 10 years higher in the top 1 percent of counties by life expectancy compared to the bottom 1 percent of counties (Dwyer-Lindgren et al. 2017). The geographic variation in life expectancy is particularly large among individuals in the lowest quartile of income (Chetty et al. 2016).

In this essay, we begin with an overview of geographic differences in life expectancy across the United States and Europe. We then discuss the problems with seeking to either infer underlying place health effects—defined as the hypothetical effect on health of relocating an individual from one location to another—or to explain the causal mechanisms behind the geographic variation with a "naïve regression" approach of regressing life expectancy on the characteristics of the area and the local population. A more promising approach that uses data on movers to identify causal effects of place on health has been developed in

■ *Tatyana Deryugina and David Molitor are both Associate Professors of Finance, Gies College of Business, University of Illinois at Urbana–Champaign, Champaign, Illinois. Deryugina and Molitor are both Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are deryugin@illinois.edu and dmolitor@illinois.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.147.

recent years, and we draw from this literature to illustrate how one might detect and measure place health effects. Finally, we discuss some possible mechanisms behind place effects.

The extent to which the observed geographic variation in life expectancy across places reflects the causal effects of place of residence cannot be observed with a simple comparison, because there are two (not mutually exclusive) ways through which such variation could arise. First, geographic variation in life expectancy could be due to non-random geographic sorting of individuals, with geographic differences reflecting the exogenous characteristics of residents. Second, place of residence could have a causal effect on longevity through a variety of channels, which may operate through largely immutable local characteristics (like climate in that area) or through characteristics amenable to policy (like the local health care system or exposure to pollution). To make matters more complicated, geographic sorting can itself give rise to place effects through peer influence on health-relevant behaviors, yielding geographic differences in life expectancy that are a product of both non-random sorting and the peer influences of individuals who live there. Understanding the contribution of each of these factors is of paramount importance for crafting optimal policy.

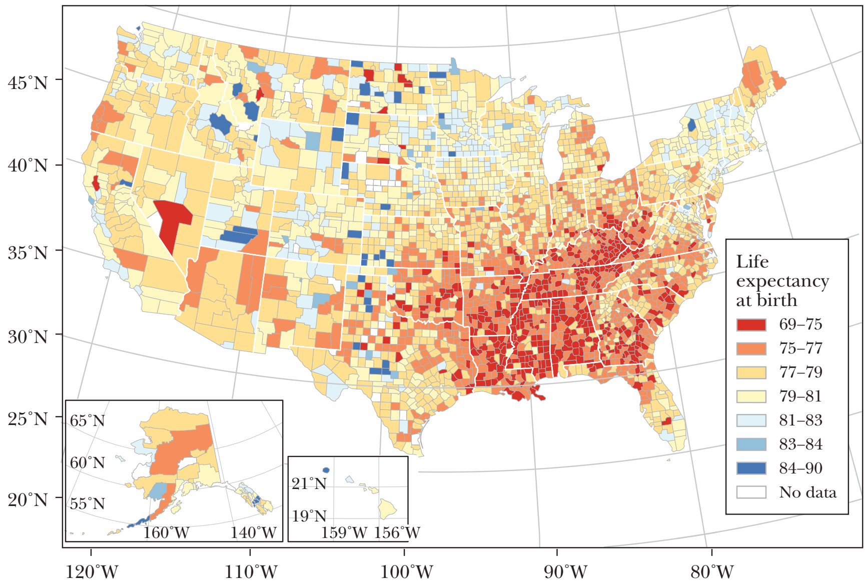
When considering how place shapes health, most of the motivation and empirical literature has focused on longevity, arguably because it is easier to measure systematically than other aspects of health. Additionally, life expectancy gains are worth a lot (Murphy and Topel 2006), making longevity a natural first-order concern for researchers and policymakers. We therefore focus most of our discussion on how place of residence affects mortality while recognizing that place could also affect other important dimensions of health.

Regional Differences in Life Expectancy: United States versus Europe

As a starting point for gauging the potential role of place in determining life expectancy, we compare how life expectancy varies across local regions in the United States and Europe. For the United States, we measure life expectancy at birth in 2010–2015 for each county using data from the US Small-Area Life Expectancy Estimates Project (Arias et al. 2018). The average population of the approximately 3,200 US counties and county equivalents is roughly 100,000, but this varies greatly from Los Angeles County with about 10 million people to the smallest counties with no more than a few hundred people.

For Europe, we measure life expectancy at birth for geographies defined by the Nomenclature of Territorial Units for Statistics (NUTS) using the NUTS 3 measure, which is the most geographically refined level in this system and most comparable to US counties. There are about 1,500 NUTS 3 regions in 37 countries across Europe, with population ranging from about 150,000 to 800,000. Life expectancy data are not systematically reported at the NUTS 3 level; we compile

Figure 1
US Life Expectancy



Source: This figure is based on our calculation of county-level life expectancy. As we describe in the online Appendix Section A.1, our calculation is based on Arias et al. (2018) and Manson et al. (2020).
Note: The figure shows county-level life expectancy at birth in the United States.

the data for 1,057 regions in 22 countries from various sources using the most recent period available for each region (Deryugina and Molitor 2021). In some analyses, we use life expectancy in 2018 at the NUTS 2 level, the next-highest level of geographic aggregation because these data are available for all European countries covered by NUTS except Albania. Online Appendix, available with this article at the *JEP* website, Sections A.1 and A.2 and Tables A.1 and A.2 describe the life expectancy data and methodology in greater detail.

As a vivid example of geographic differences in mortality across the United States, Fuchs (1974) compared mortality rates in Nevada and Utah, which are neighboring states with similar climates and, at the time, similar income levels and physicians per capita. Fuchs noted that, nonetheless, adult mortality rates were substantially higher in Nevada than in Utah, which he attributed to Nevada's high rates of cigarette and alcohol consumption as well as "marital and geographical instability." Even today, the average person born in Utah has a life expectancy 1.9 years higher than the average person born in Nevada.

More generally, Figure 1 and the first column of Table 1 reveal that life expectancy at birth varies widely across US counties, from a low of 69.1 years in East Carroll Parish in northeastern Louisiana to a high of 89.5 years in Cheyenne

Table 1
Regional Life Expectancy in the United States and Europe

	United States	*Europe*
Geographies		
Regional unit of analysis	County	NUTS 3
Number of regions	3,108	1,057
Number of states and District of Columbia (US only)	51	
Number of countries	1	22
Life expectancy at birth		
Period	2010–2015	2011–2019
Mean	77.7	80.6
Standard deviation	2.6	2.5
Minimum	69.1	70.4
10th percentile	74.4	76.7
90th percentile	81.0	83.2
Maximum	89.5	84.8

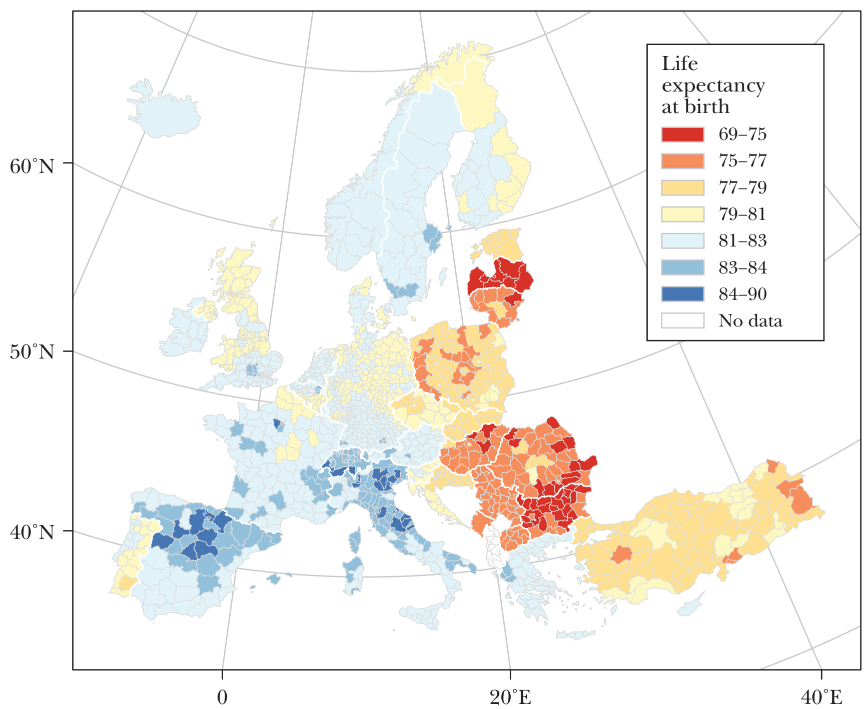
Note: The table reports summary statistics for the US and European life expectancy data samples. Details of calculations available in the online Appendix.

County, Colorado.[1] County-level life expectancy averages 77.7 years and has a standard deviation of 2.6 years. The top 10 percent of counties have a life expectancy of 81.0 years or more, while the bottom 10 percent have a life expectancy of 74.4 years or less, resulting in an interdecile range of 6.5 years. Although there is some spatial correlation in life expectancy—counties with the lowest life expectancy tend to be in the South—roughly half of the variation in life expectancy across all counties occurs between states, with the other half occurring across counties within states.[2]

Figure 2 and the second column of Table 1 show the geographic distribution in life expectancy in Europe, which ranges from a low of 70.4 years in Latgale, Latvia, to a high of 84.8 years in a region that makes up a portion of Madrid, Spain. Average life expectancy in the NUTS 3 regions in our sample is 80.6 years, with a standard deviation of 2.5 years. Like the United States, the interdecile range is 6.5 years: the top 10 percent of NUTS 3 regions have a life expectancy of 83.2 years or more, while the bottom 10 percent have a life expectancy of 76.7 years or less. Unlike the United States, approximately 87 percent of the variation in life expectancy across Europe can be accounted for by between-country variation rather than within-country variation.[3]

[1] Online Appendix Table A.3 lists the top and bottom 10 counties, by life expectancy.
[2] The R^2 from regressing county-level life expectancy on state fixed effects is 0.46, revealing that just over half (54 percent) of the variation in life expectancy across counties occurs within states, with the remainder occurring across states. See online Appendix Section A.3 for details of this regression.
[3] The R^2 values from a regression of NUTS 3 or NUTS 2 level life expectancy on country fixed effects are 0.87 and 0.85, respectively, which reveals that 85–87 percent of the variation in life expectancy across European regions is explained by the country of residence. See online Appendix Section A.3 for details of this regression.

Figure 2
European Life Expectancy



Source: This figure is based on our calculation of NUTS 2 and NUTS 3 level life expectancy. Our calculations are based on data from a variety of sources, described in online Appendix Section A.2.
Note: The figure shows life expectancy at birth in Europe. Gray borders delineate NUTS 3 regions. Life expectancy is shown at the NUTS 3 level whenever available and at the NUTS 2 level otherwise. Online Appendix Figures A.1 and A.2 show separate maps for the NUTS 2 and NUTS 3 samples, respectively.

Three main results emerge from comparing the regional variation in life expectancy in the United States and Europe. First, average life expectancy is 2.8 years higher in Europe than in the United States. Second, the overall variation in life expectancy, as captured by the standard deviation or interdecile range of the life expectancy distribution, is similar in both contexts. Third, most of the regional variation in life expectancy in Europe is explained by country of residence, whereas in the United States, most of the variation is within-state.

The reasons behind the large differences in the spatial correlation in local-area life expectancy between Europe and the United States are not immediately clear. American states are arguably more similar to each other in terms of policies than are European countries. Thus, one might expect country-level place health effects could be more heterogeneous in Europe compared to state-level place health effects in the United States, while within-country place health effects could be less heterogeneous than within-state. However, population sorting and individual preferences could also be much more heterogeneous across European countries than across American states. Thus, the life expectancy patterns seen in Europe and the United States do not necessarily help rule out or support any particular explanation. However, the geographic variation in European life expectancy—which, to our knowledge, has not been comprehensively documented at such a granular spatial level until now—demonstrates that large regional differences in life expectancy are not just a US phenomenon.

A Naïve Regression Approach

One approach to investigating whether place affects life expectancy, and if so, how, is to regress local life expectancy on the characteristics of the area and the population. We conduct such an exercise here, similar in approach to Dwyer-Lindgren et al. (2017) except we consider a somewhat different set of local characteristics and weight regressions by each county's population, as life expectancy is likely to be measured with greater error in smaller counties.[4] Because a simple regression of local life expectancy on local area and population characteristics cannot account for a number of important confounders, we dub it a "naïve regression approach."

In Table 2, panel A shows the results of bivariate regressions with US county-level life expectancy as the dependent variable and a variety of local health and environmental characteristics as the explanatory variables. Life expectancy is positively correlated with the percent of population that exercises and is negatively correlated with smoking and obesity rates. The local smoking rate alone explains over 46 percent of the cross-sectional variation in life expectancy, as indicated by the R^2. Obesity and exercise rates individually explain about 42 and 34 percent of the variation, respectively.

Health care quantity—as measured by the number of doctors per capita and the number of hospital beds per capita—each explain 6.0 and 4.4 percent of local life expectancy, respectively. The correlation between life expectancy and the number of hospital beds per capita, however, does not have the expected sign: more hospital beds is associated with *lower* life expectancy. As we discuss further below, this counterintuitive correlation hints at the difficulties inherent in recovering the mechanisms behind place health effects. In this case, for example, more hospital beds could be a response to poor health and elevated health care needs among residents.

Finally, local environmental quality, as measured by fine particulate matter ($PM_{2.5}$) concentrations, explains almost 5 percent of the geographic variation in life expectancy. Climate, as measured by the number of hot (>90°F) days per year,

[4] We focus on the local characteristics considered by Deryugina and Molitor (2020), except those derived from Medicare claims. See online Appendix Section A.4 for data and regression details and Deryugina and Molitor (2021) for the data. Online Appendix Table A.5 shows the results of unweighted regressions.

Table 2
The Relationship between County-Level Life Expectancy and Local Characteristics

County characteristic	*Mean [standard deviation]*	*Ordinary least squares coefficient (standard error)*	R^2
A: Health and environmental characteristics			
Percent smoking	21.29 [4.05]	–0.36 (0.02)	0.462
Percent obese	20.10 [4.12]	–034 (0.02)	0.423
Percent exercising	74.74 [5.44]	0.23 (0.02)	0.341
Physicians per 1,000 capita	2.77 [1.94]	0.28 (0.05)	0.060
$PM_{2.5}$ concentrations	10.38 [1.94]	–0.24 (0.08)	0.049
Hospital beds per 1,000 capita	3.40 [2.55]	–0.18 (0.03)	0.044
Hot days/year (90°F+)	2.21 [8.59]	–0.01 (0.01)	0.001
	0.78 [0.06]	0.95 (1.71)	0.001
B: Economic characteristics			
Median home values ($1,000s)	128.87 [65.91]	0.02 (0.00)	0.490
Income per capita ($1,000s)	21.63 [5.28]	0.24 (0.01)	0.344
Poverty rate, 65+	0.10 [0.04]	–20.97 (3.36)	0.181
Upward income mobility (from p25)	–0.03 [0.41]	1.88 (0.34)	0124
Urban population share	0.79 [0.25]	2.47 (0.33)	0.081
Crime rate per 1,000	7.62 [3.49]	–0.14 (0.03)	0.052
Local gov. spending per capita	2.51 [1.06]	0.46 (0.15)	0.050
Upward income mobility (from p75)	–0.03 [0.23]	–1.63 (0.60)	0.029
Social capital index	–0.46 [1.11]	0.25 (0.11)	0.016
Income segregation	0.07 [0.03]	6.61 (2.87)	0.011
C: Multivariate comparisons			
All health and environmental characteristics			0.695
All economic characteristics			0.672
All characteristics			0.807

Note: The table reports results from regressing US county-level life expectancy on local characteristics. Each row in the table corresponds to a separate regression, where the included local characteristic(s) are indicated by the row labels. Observations are weighted by county population. Column 1 shows the mean and standard deviation (in brackets) of the local characteristic. Column 2 reports regression coefficients and robust standard errors (in parentheses). Column 3 reports the R^2 from the regression. Online Appendix Section A.4 provides more details on the data and regressions.

explains only about 0.1 percent of the geographic variation and has no statistically significant relationship with local life expectancy.

Panel B also considers bivariate regressions, this time using economic characteristics as the explanatory variables. Median home values explain almost one-half of the geographic variation in life expectancy, and local income per capita explains more than one-third. The elderly poverty rate explains about 18 percent of the geographic variation, and upward income mobility from the 25th percentile explains about 12 percent. The share of the population living in an urban area explains about 8 percent of the variation, and per-capita spending by the local government and the local crime rate each explain about 5 percent. Finally, upward income mobility from the 75th percentile and income segregation explain 2.9 and 1.1 percent of the geographic variation, respectively.

Most of the characteristics mentioned above are significantly correlated with each other. In panel C, we report the R^2 values from regressing life expectancy on bundles of characteristics. All the health and environmental characteristics combined explain 69.5 percent of the cross-sectional variation in life expectancy, while all the economic characteristics explain 67.2 percent. The most complete regression that includes all the characteristics mentioned above explains 80.7 percent of the variation in life expectancy. These results are qualitatively similar to those of Dwyer-Lindgren et al. (2017), who use a slightly different set of health and socioeconomic characteristics to conclude that these characteristics explain as much as 74 percent of the unweighted county-level variation in life expectancy.

A key problem with concluding from the results in Table 2 that place of residence has a causal effect on health is that people with different health prospects and behaviors may endogenously sort into different locations. As emphasized by Roback (1982), individuals choose where to live based in part on the amenities in each region. If people who are predisposed to good health place a higher value than others on amenities that will extend their life expectancy, such as low levels of air pollution, then regional health differences arise partly because of sorting, and observed health differences will overstate the causal impacts of place. By contrast, if relatively unhealthy individuals place a high value on such amenities, regional differences in health will understate the causal role of place.

One possible approach to account for population sorting is to control for individual characteristics and behavior to see how much of the regional difference in health remain unexplained. Early literature investigating place health effects did exactly this, attributing the residual variation to place-specific health effects (for an overview, see Macintyre, Ellaway, and Cummins 2002). However, this approach to measuring place health effects will not yield correct estimates without extremely restrictive assumptions, such as that certain characteristics reflect only sorting while others only capture the causal effect of place. Yet individual behaviors, characteristics like education and income, and even the demographic composition in a place could reflect the effects of living in that place (Macintyre, Ellaway, and Cummins 2002). As a result, health differences that seem to be explained by individual characteristics could also arise through the causal effects of place on choices, behavior, and aging. Furthermore, if health is transmitted intergenerationally, observed geographic differences in health could reflect sorting of individuals' ancestors, further complicating estimation.

Given the abovementioned issues, what do we learn from naïve regressions like those reported in Table 2? On the one hand, one might argue that many of the correlations between local characteristics and life expectancy reflect place effects, perhaps via peer effects or amenities like a local economy that causes incomes to be higher or lower. On the other hand, if differences in socioeconomic status and behavioral patterns are not *caused* by place of residence, referring to them as "place effects" is a misnomer. For example, these differences could be due partly to sorting and population movement, including from a long time ago, creating cross-county differences that have less to do with place of residence and more to do with genetic predisposition and upbringing not directly related to the geographic location. In this case, exogenously moving individuals to a particular location would not alter their health, and therefore these kinds of geographic differences cannot be considered true "place effects."

This list of concerns about the interpretation of regression results can easily be expanded. Ultimately, one might just conclude that correlations are not causation, and no reliable lessons about place effects of health can emerge from a naïve regression approach.

Using Movers to Identify Causal Effects of Place

Gauging the Magnitude of Place Health Effects

Empirical difficulties notwithstanding, economic theory predicts that place health effects are likely to exist. For example, a spatial sorting model in the style of Roback (1982) posits that places differ in their local amenities, such as the climate, pollution levels, and the quantity and type of leisure opportunities. Some amenities, like a mild climate or a good harbor, facilitate certain types of firm production in that location; other amenities, like clean air assured through local government anti-pollution regulations, will increase the cost of production and decrease output, all else equal. In the most basic spatial sorting models, individuals can move costlessly and choose where to live based on local wages, rents, and amenity levels. In equilibrium, utility is equalized across places, but differences in the quantity and productivity of local amenities give rise to spatial heterogeneity in wages and rents. The resulting variation in amenities and disposable income, in turn, implies that longevity is likely to be influenced by one's place of residence. But are the health effects of place generated in this way likely to be large in magnitude?

Gauging the magnitude of place health effects is difficult for several reasons. Accounting for the distinct role of sorting in local life expectancy is challenging, not least because individual-level sorting based on an area's amenities may (directly or indirectly) give rise to place health effects and do so in offsetting ways. For example, while regions with high amenities for certain kinds of production will have higher wages, sorting of people into that area will raise rents in those regions (and perhaps suppress real wages for workers in other industries), potentially counteracting the direct effects of higher incomes on health to some extent. In practice, the direct and indirect effects are unlikely to offset fully in all places, especially because people do not value all amenities solely for their effects on longevity. A related challenge is that there may be peer effects in health-relevant behaviors like smoking or exercise that exacerbate the impact of any population sorting on the equilibrium levels of spatial heterogeneity in life expectancy. Finally, heterogeneity in local amenities as well as heterogeneity in individuals' preferences, productivity, or information can also yield heterogeneity in place health effects. If someone likes a place for its hiking trails and someone else likes it for its lively nightlife, for example, the causal effects of that place on the life expectancy of these two individuals may be of opposite signs.

Thus, even the simplest sorting models predict the existence of place health effects but also make clear that measuring them empirically is challenging. Of course, the strict assumptions of such models are not likely to be satisfied: people do not have full information about health (or other) prospects of potential locations, nor is it costless to move. But the reality of differences in amenities across locations, and the likelihood that these differences will in some way cause differences in health, remains.

As a conceptual experiment, place health effects could be quantified and disentangled from sorting effects by randomly assigning individuals to different places of residence and measuring differences in their subsequent life expectancy. While conducting such an experiment is impractical for many reasons, a related approach of looking at movers coming from the same place and ending up in different locations, for example, can help separate the causal effects of place from various confounders. To our knowledge, all quasi-experimental papers that speak to the causal effect of place on health leverage movers in some form. An earlier literature compares the health of movers to non-movers, while a more recent one compares different groups of movers.

Spatial equilibrium models caution against a possible pitfall from using movers to identify the causal effects of place on health: if individuals sort into locations in equilibrium, then the movers may be no less selected than individuals who already reside in a particular location. However, an advantage of a design that uses movers is that, as long as movers are observed for a reasonable period of time before a move, such sorting can generally be evaluated and potentially accounted for. We next discuss the research designs of quasi-experimental studies of place health effects in more detail and summarize the conclusions the literature has reached thus far.

Comparing Movers to Non-Movers

In some studies of health effects related to moving, researchers have sought to address identification problems by looking for factors that are predictive of certain individuals moving, but plausibly exogenous with respect to future health. If these two conditions are satisfied, the predictors of moving can be used either as instruments or in a reduced-form way to estimate the causal effects of moving on health. For example, Gibson et al. (2013) exploit a migration lottery to estimate the causal effect of migration from Tonga to New Zealand on blood pressure and the prevalence of hypertension. In a study of the long-run mortality effects of the early twentieth-century Great Migration of African Americans from mostly rural locations in the Deep South to mostly urban locations in the North, Black et al. (2015) use the proximity of individuals' birthplaces to railroad lines as an instrument for migration. Johnson and Taylor (2019) build on this identification strategy and use the timing of railroad construction as well as patterns of postal mail flows to estimate the mortality effects of the mid-twentieth-century migration from rural locations in the Northern Great Plains states to urban locations in the American West and Midwest.

These earlier studies focus on estimating the health effects of migration rather than of place, so their estimates will reflect both the health effects of the act of migrating itself (for example, due to losses of community ties) as well as the average effect on health of living in the destination regions. More generally, these studies are unable to pin down the exact mechanisms through which migration affects longer-run health. For example, they cannot determine the extent to which the specific composition of origin and destination regions matters for the estimated effects: for example, the origins are mostly rural and destinations mostly urban in the studies of Black et al. (2015) and Johnson and Taylor (2019).

Nonetheless, these studies provide suggestive evidence that health effects of place exist and are nontrivial in magnitude. Gibson et al. (2013) estimate that Tonga-to-New-Zealand migration raises blood pressure and increases hypertension prevalence by 11 percentage points or about one-third of the mean among lottery losers. Black et al. (2015) find that, conditional on surviving to age 65, leaving the Deep South lowered life expectancy by at least 1.5 years. They also show that the movers smoked and drank significantly more than those who did not migrate. Correspondingly, movers to the North are substantially more likely to die from respiratory cancer, chronic obstructive pulmonary disease, and chronic liver disease and cirrhosis. Likewise, Johnson and Taylor (2019) find that the mid-twentieth-century US migration from rural to urban areas increased mortality and provide suggestive evidence that this is due to increased smoking and alcohol consumption.

Other indirect evidence that local conditions matter for health comes from papers that use movers to study how local conditions affect health care provision and other non-health outcomes that could ultimately affect health. For example, Song et al. (2010) show that when Medicare recipients move between regions, rates of medical diagnoses change. Finkelstein, Gentzkow, and Williams (2016) study Medicare recipients who move between areas and show that place of residence affects movers' medical spending. Molitor (2018) looks at cardiologists who move and finds that, on average, their own practice patterns change by 60–80 percent of the difference in local norms between their new and original practice regions. Mover designs have also shown that local conditions can affect levels of education and earnings (Chetty, Hendren, and Katz 2016; Nakamura, Sigurdsson, and Steinsson 2017; Chyn 2018; Chetty and Hendren 2018). To the extent that each of these factors matters for health, we might therefore surmise that place of residence will have health consequences via such channels. As noted earlier, however, such analysis would need to account, for example, for both direct and indirect health effects of living in a higher-income and higher-cost-of-living area.

Comparing Movers to Other Movers

Studies of place health effects that do not use movers cannot thoroughly assess the degree of sorting into a location and therefore cannot control for it without restrictive assumptions (like assuming that sorting only operates through immutable characteristics like race or age). Studies that compare movers to non-movers can make progress on this dimension, but nonetheless cannot separate place health effects from the health effects of moving itself.

A substantially more credible research design for estimating place health effects comprises comparing movers to each other, something that several recent studies have done. This research design is based on the insight that if two otherwise identical individuals initially living in the same place simultaneously move to different destinations, then subsequent differences in their health will be due to place effects. Medicare recipients who move between counties offer a promising source of evidence in this area, in part because they can maintain their health insurance coverage when they move. Additionally, Medicare administrative data include the vast majority of US elderly and long-term disabled individuals and provide detailed health utilization records for many of them, allowing researchers to control for differences in observable characteristics before the move and to assess the extent of non-random sorting into destination regions. However, researchers must still overcome the challenge that individuals who move to different destination regions may not be identical and their destination choices may not be exogenous with respect to other unobserved determinants of health. This hurdle has proven formidable, and the literature is still in its nascency.

How can researchers overcome the difficulty of movers sorting non-randomly into destinations? Statistical identification of place health effects in studies that compare movers to each other does not require that movers choose their destination region completely at random. Instead, identification requires that, conditional on available controls, movers' choice of destination is unrelated to any other future determinants of the health outcome of interest. Studies interested in relating health outcomes to some specific characteristic of place—such as the local mortality or obesity rate—require an even weaker identification assumption to interpret that correlation as proxying for the causal effect of place on health: the destination characteristic of interest must be unrelated to unobserved determinants of future health. For example, movers selecting destinations based on whether they have relatives living there does not confound the research design as long as either the presence of relatives *or* the health effects of living near relatives are orthogonal to the destination characteristic being used as the proxy for place health effects. As we discuss in the next section, however, while a correlation between a destination characteristic and changes in movers' health can be interpreted as demonstrating that place has a causal effect on health if the abovementioned assumption holds, the relationship cannot be interpreted as the causal effect of *that particular characteristic* without additional assumptions.

Directly testing the identification assumptions discussed above is infeasible because one can never be sure they are observing all determinants of future health. However, one indirect test involves estimating whether a destination characteristic of interest is correlated with preexisting trends in the health outcome(s) of interest. Research using outcomes other than mortality can assess the likelihood of such endogenous sorting by explicitly estimating such trends among movers before the move. For example, Baum et al. (2020) use administrative records from the Veterans Health Administration to study how a mover's probability of having uncontrolled chronic conditions (hypertension, diabetes, obesity, or depression) is affected by the local prevalence of such conditions. The authors show that movers to regions that differ in the prevalence of uncontrolled chronic conditions do not exhibit differential pre-trends in such conditions before the move. Thereafter, moving to a ZIP code with a greater prevalence of a given chronic condition increases the probability of being diagnosed with that condition within three years of the move. The magnitude of the estimated effects varies from 3.1 percent of the change in the local prevalence for obesity to 27.5 percent of the change in the local prevalence for hypertension.

Because one must be alive to move, a direct test of parallel trends in mortality before a move is not possible, and other approaches to assess and control for any differential sorting must be used. Assessing sorting by using *predictors* of mortality is one such approach. Deryugina and Molitor (2020) study how the mortality of Medicare beneficiaries displaced by Hurricane Katrina relates to mortality in their destination county. They construct an index of predicted mortality for each mover from extensive measures of chronic conditions and spending histories and show that movers' predicted mortality is uncorrelated with mortality in their destination. There is, however, an almost one-for-one relationship between movers' realized mortality and the mortality of residents in their destination county, which suggests that where movers relocated had a causal effect on their longevity.

A sophisticated approach to control for sorting is developed by Finkelstein, Gentzkow, and Williams (2021), who use the relocation of Medicare beneficiaries to estimate the causal effects of place on mortality. Their definition of place consists of "commuting zones," which are aggregations of counties chosen to approximate local labor markets. (In 2000, the United States had 709 commuting zones.) The authors control for sorting using a generalization of the method developed by Oster (2019), which uses variation in an observable variable (in their study, the correlation between choice of destination and observed health characteristics) to adjust for variation in an unobservable variable (in their study, the correlation between choice of destination and unobserved health characteristics). Finkelstein, Gentzkow, and Williams can estimate place-specific mortality effects and can therefore directly study heterogeneity in place health effects, unlike Baum et al. (2020) and Deryugina and Molitor (2020). Finkelstein, Gentzkow, and Williams estimate that equalizing health-related place effects across US commuting zones would reduce the geographic variation in life expectancy of 65-year-olds by 15 percent.

As discussed earlier, place effects may be heterogeneous across places and across individuals. Consistent with this prediction, Finkelstein, Gentzkow, and Williams (2021) show that estimated place effects on longevity vary widely across the United States. Thus, different research designs may arrive at varying conclusions because of differences in where the in-sample movers are relocating. It is also possible that the health effects of a given place vary across individuals themselves: for example, Chetty et al. (2016) find that the largest regional health disparities occur among the poorest individuals (those in the bottom 5 percent of income), suggesting place matters more for this group. The presence of heterogeneous place effects implies that who the marginal mover is could affect a study's estimates.

Given the small number of studies leveraging movers to estimate the causal effects of place on health and their somewhat heterogeneous methods, a direct comparison of most existing results is difficult. Finkelstein, Gentzkow, and Williams (2021) relate the place effects they estimate to local average life expectancy and find a positive correlation: moving to a place where life expectancy is one year higher causes the mover to live 0.23 years longer, on average. This relationship is substantially smaller than the almost one-for-one relationship estimated by Deryugina and Molitor (2020). If place effects are heterogeneous, which seems likely, then the differences between these two studies could be due to differences in destination regions or in the composition of movers. For example, Hurricane Katrina almost certainly displaced many people who would otherwise not have moved, whereas the movers exploited by Finkelstein, Gentzkow, and Williams are typical elderly movers. Of course, it is also possible that one or both studies failed to account properly for non-random sorting into destination regions.

While recent research using state-of-the-art movers' design points to a causal relationship between place of residence and health and longevity, these findings must be verified and extended. Both Finkelstein, Gentzkow, and Williams (2021) and Deryugina and Molitor (2020) use Medicare data, and therefore their sample of movers consists of older individuals and, in the case of Deryugina and Molitor (2020), the long-term disabled. Baum et al. (2020) has a sample of US veterans that is younger, on average, but is overwhelmingly male. The extent to which place of residence matters for the health of younger individuals, especially younger women, therefore remains an important question for future research.

Another important shortcoming of the papers estimating place health effects is that none speak to the welfare impacts of migration, which could differ qualitatively from estimated health effects both because there are costs to moving and because there can be benefits of living in a place other than its effects on health. Whether the observed migration was welfare-improving on net may vary by context. In the case of Finkelstein, Gentzkow, and Williams (2021) and Baum et al. (2020), for example, migration is likely voluntary and thus plausibly welfare-improving. By contrast, many movers studied by Deryugina and Molitor (2020) were forced to move by Hurricane Katrina and may have suffered a welfare reduction on net. But there exists no direct evidence on whether encouraging migration to places with favorable health effects would be welfare-improving.

Channels through which Place May Affect Health and Longevity

Our discussion has already hinted at the various channels through which place may affect health. We now consider them systematically and in more detail. Understanding these channels will inform whether or how policies can be designed to improve population health. For example, policy implications are rather different if place health effects are driven by immutable local characteristics, such as climate, compared to if they are driven by peer effects or by public policies.

The ideal experiment to understand the mechanisms behind measured place health effects involves exploiting an exogenous change in some local characteristic and estimating the subsequent change in local life expectancy. Such experiments—whether natural or implemented by a researcher—are rare. Estimating the presence and magnitudes of specific mechanisms has thus proven difficult, and the current evidence in this area is mostly suggestive. In addition to the identification assumptions required to establish that place of residence has a causal effect on health, establishing a causal relationship between any specific local characteristic and health in an observational or experimental setting requires an additional assumption: the local characteristic must not be correlated with any other unobserved local determinant of health. Given the variety of local characteristics that may matter for health, it is unlikely that this assumption is valid for any existing study of place health effects.

For example, greater economic activity could both raise residents' incomes and increase air pollution. Even if one can establish that living in that particular area raises life expectancy on net, separating the contribution of higher income from that of higher air pollution is challenging because both are generated by difficult-to-quantify "economic activity." A naïve regression of life expectancy on local air pollution may even yield counterintuitive positive correlations. Such difficulties are not limited to cross-sectional studies: Deryugina and Molitor (2020) indeed find that higher local concentrations of $PM_{2.5}$ are associated with lower mover mortality. Similarly, Finkelstein, Gentzkow, and Williams (2021) estimate that places that are good for longevity tend to have fewer hospital beds per capita.

Identification challenges notwithstanding, a variety of studies that do and do not exploit movers have examined how local characteristics correlate with life expectancy and health—and have largely been careful not to interpret them as causal. In part due to statistical power considerations and high degrees of correlation between some local characteristics, most research that uses movers has considered local characteristics separately, rather than jointly. Thus, just as estimated place health effects potentially reflect the influence of a bundle of characteristics, the specific local characteristic(s) identified as predictive of place health effects could be proxies for the influence of a group of correlated characteristics.

Some insight about the mechanisms behind place health effects can also be gleaned from exploiting plausibly exogenous region-wide changes in policy, such as smoking or health care regulations. Yet another approach is to use experimental or quasi-experimental methods to study the causal health effects of factors that vary across individuals rather than regions. Numerous such studies exist. While they may be indirectly informative about the mechanisms that could generate the observed place effects, they cannot speak to place effects directly because of their piecemeal approach. After all, the regional distribution of positive and negative contributors to life expectancy could be such that some factors counteract each other and only a few are important for explaining place effects on aggregate.

There are five broad, interrelated mechanisms that could be generating observed place health effects: socioeconomic status, peer effects, health care delivery, the local environment, and public policy. They are interrelated because of their potential to influence each other; for example, elevating the socioeconomic status of some of an area's residents could affect others through peer effects or through greater demand for pollution reduction. There may also be peer effects among health care providers, which influences the types and intensity of health care provided in different areas (for example, Molitor 2018). We discuss each of these mechanisms in turn and summarize the available evidence on their importance, drawing both from studies that do and do not use movers.

The first mechanism is socioeconomic status, such as income and wealth, employment conditions, and education. While these channels are unlikely to explain the place effects among elderly and disabled individuals, such as those studied by Deryugina and Molitor (2020) and Finkelstein, Gentzkow, and Williams (2021), they may be important components of place health effects for younger and working-age individuals. If income affects health and if places estimated to be good for health also tend to increase younger movers' income and employment, then estimates of place effects based on moves later in life will capture only part of the overall effect of place on longevity. More generally, the effects of place later in life may correspond less than one-to-one with regional mortality outcomes, even if all regional differences in mortality are due to place of residence. An opportunity for future research is therefore to measure how place of residence earlier in life matters for health and longevity.

A considerable body of evidence not directly related to place effects does suggest that socioeconomic status plays a key role in building and maintaining health (Grossman 1972). For example, Frijters, Haisken-DeNew, and Shields (2005) and Lindahl (2005) find that plausibly exogenous income shocks improve self-reported health, and Schwandt (2018) finds that negative wealth shocks due to stock market fluctuations impair physical and mental health and increase mortality. Job separations have been linked to elevated mortality risk for decades post-separation (Sullivan and Von Wachter 2009), and young people first entering the labor market during a recession face higher mortality risks later in life (Schwandt and Von Wachter 2020). The level and quality of education individuals receive can also influence mortality (Buckles et al. 2016; Galama, Lleras-Muney, and van Kippersluis 2018).

Even though the movers in their sample are largely not in the labor force and have completed their formal education, both Deryugina and Molitor (2020) and Finkelstein, Gentzkow, and Williams (2021) find that moving to areas with higher socioeconomic status is beneficial for survival. Although these findings could reflect the influence of other local characteristics that are simply correlated with socioeconomic status, living in a higher socioeconomic status area might also offer indirect health benefits if higher socioeconomic status causes such areas to develop amenities that are beneficial for health. For example, proximity to grocery stores or restaurants with nutritious food may facilitate healthy living, and areas with higher socioeconomic status may attract more such establishments. The empirical evidence on this specific mechanism is mixed. Allcott et al. (2019) find that the nutritional quality of purchased groceries is not affected by moves to neighborhoods with greater availability of healthy groceries, and Hut (2020) finds no relationship between average nutrition quality of purchased groceries in a destination and changes in movers' nutritional quality for at least two years following the move. But Currie et al. (2010) find that the presence of a fast-food restaurant near a school raises the probability of obesity among the students. If a similar dynamic operates for adults and if demand for fast food is lower in areas with higher income or education, then the presence of healthier restaurant foods may be one mechanism behind place effects.

The second potential mechanism behind place health effects is peer effects. Moving can change one's peers and in this way give rise to peer effects in health behaviors, ultimately affecting a mover's health. Studies in this domain face well-known identification challenges (Manski 1993; Angrist 2014). Most of the research regarding peer effects on health-related behaviors has been done on students and young people and none has been directly related to place effects. Sacerdote (2001) finds that a randomly assigned college dormitory roommate's drinking behavior does not influence one's own drinking behavior, but overall drinking behavior by dormmates does, suggesting the existence of peer effects at higher levels on this dimension. Of course, such peer effects may look very different outside of a college dormitory. Fletcher (2010) combines an instrumental variables strategy with fixed effects to show that classmates' smoking behavior affects one's own. Card and Giuliano (2013) also find peer effects in smoking among youths. Angrist (2014) notes that the best-identified studies have largely found effects small in magnitude, but this, of course, does not rule out their existence in the context of place effects. Additionally, there may be peer effects along other health-relevant dimensions, such as preventive care utilization or regular health screenings, for which there is virtually no well-identified empirical evidence. Overall, whether peer effects are present among older adults and whether they are large enough to generate meaningful differences in health behaviors on aggregate remains an open question.

Both studies of rural–urban movers discussed earlier conclude that movers increase their consumption of alcohol and tobacco, potentially explaining their decrease in life expectancy (Black et al. 2015; Johnson and Taylor 2019). Baum et al. (2020) find that post-move changes in the prevalence of uncontrolled hypertension, obesity, diabetes, and depression are each significantly correlated with the destination region's prevalence of the same condition.[5] These correlations are consistent with peer effects, but other explanations are possible. Among rural–urban movers, it could be that alcohol and tobacco consumption increased because of higher incomes after moving to the city rather than driven by peer effects. Among movers more generally, health habits of both movers and local residents could simultaneously be affected by a variety of living conditions, including local prices and policy, giving rise to the observed correlations.

[5] Deryugina and Molitor (2020) find that rates of smoking, obesity, and exercise at the destination location are significantly associated with movers' subsequent mortality. Similarly, Finkelstein, Gentzkow, and Williams (2021) find a positive relationship between the effect of a place on mortality and smoking and obesity rates and a negative relationship between place effects and exercise rates.

A third potential mechanism is the quality or quantity of health care delivery, which could have both short- and long-term health effects on residents of all ages. Per-capita health care spending varies substantially across US regions, making it at least a plausible determinant of local life expectancy. Correlational studies have found that, on average, regions with higher levels of per-capita health spending have no better health outcomes than lower-spending regions (for a discussion, see Skinner and Fisher 2010). Finkelstein, Gentzkow, and Williams (2021) find that positive place effects are correlated with higher quality and quantity of health care, but Deryugina and Molitor (2020) find no relationship between movers' mortality and local medical spending or health care quality. The potential for reverse causality and other possible confounders makes it difficult to conclude that either of these correlations are causal. Other studies using quasi-experimental evidence suggest that, at least in some settings, higher health care spending is beneficial for health. Using different identification strategies, Doyle (2011) and Doyle et al. (2015) find that patients randomly hospitalized in a higher-spending region or hospital, respectively, are less likely to die. Doyle, Graves, and Gruber (2017) confirm these findings for inpatient spending, but also show that higher outpatient spending by hospitals is associated with lower survival of patients who are randomly transported there.

Health care access also varies geographically in the United States and could have meaningful impacts on health. Finkelstein et al. (2012) find that randomly selected recipients of Medicaid in Oregon reported better physical and mental health after a year with health insurance, but find no clinical evidence of better health. Miller, Johnson, and Wherry (2019) show that mortality among the near-elderly fell by almost 10 percent in states that participated in the Affordable Care Act Medicaid expansion, compared to states that did not. Abaluck et al. (2020) find that, conditional on being insured, specific health insurance plans affect beneficiaries' mortality rates. Combined with geographic differences in plans' availability, this study suggests another possible mechanism behind place effects.

The fourth potential mechanism behind observed place health effects is environmental quality, which varies considerably across the United States. Numerous studies have shown that air pollution has a causal effect on both infant and older adult mortality.[6] Similarly, both abnormally cold and abnormally hot temperatures have been shown to increase the mortality rates of the elderly (Barreca et al. 2016; Deschênes and Greenstone 2011; Heutel, Miller, and Molitor 2020). Finkelstein, Gentzkow, and Williams (2021) find that positive place effects are correlated with lower pollution levels and a more moderate climate. However, Heutel, Miller, and Molitor (2020) show that places adapt to common temperature, whether hot or cold, implying that even though abnormal temperatures raise mortality, the total contribution of the local temperature climate to regional differences in life expectancy may be small.

[6] See, for example, Chay and Greenstone (2003a, b), Currie and Neidell (2005), Currie and Walker (2011), Chen et al. (2013), Knittel, Miller, and Sanders (2016), Barreca, Neidell, and Sanders (2017), Deschênes, Greenstone, and Shapiro (2017), Ebenstein et al. (2017), Deryugina et al. (2019).

The fifth mechanism is public policy, which can affect everything from socioeconomic status to air pollution levels. Policy can also affect life expectancy through many channels not discussed above, such as by influencing individuals' smoking and drinking behavior directly or by providing a variety of social safety nets to low-income households. The most direct evidence on the role of policy in influencing the geographic variation in life expectancy comes from Montez et al. (2020), who relate state-level changes in life expectancy to changes in policy over the period 1970–2014. They examine 135 different policies, categorizing each as liberal (defined as increased regulation of the economy by the state or increased protection of marginalized groups) or conservative and then creating 18 time-varying policy indices that group related policies together. The authors find that more liberal policies tend to be associated with improved life expectancy for both men and women and that this relationship is particularly strong for the regulation of private labor, immigration, civil rights, and the environment. Higher tobacco taxes are also associated with increased state-level life expectancy over this time period.

While research on the mechanisms behind place effects has produced some suggestive correlations, the fact that some local characteristics have consistently been shown to be correlated with the life expectancy of movers and non-movers alike does not imply that the literature has successfully identified the mechanisms behind measured place effects because of the likelihood of unobserved confounders. Related areas of research offer stronger evidence on several potential mechanisms, such as income and health care access, but cannot be used to quantify the magnitude of place effects without restrictive assumptions.

Further complicating the study of mechanisms is that their health effects could be heterogeneous. Consistent with this idea, Chetty et al. (2016) document that the standard deviation of life expectancy across commuting zones is 1.4 years for men in the bottom income quartile but only 0.70 years for men in the top income quartile. While they do find that some of the important correlates of life expectancy are similar for low- and high-income individuals—such as smoking, exercise, and obesity rates—correlations between life expectancy and other local characteristics are sometimes significantly different for these two groups. Similarly, Montez et al. (2019) find that, conditional on birth state and basic demographics, there is little variation in state-level life expectancy for those with at least one year of college but a substantial amount of variation for those without a high school degree. Due to the difficulties with interpreting cross-sectional analyses causally, such patterns of course do not prove that place effects and mechanisms are heterogeneous, but they do provide suggestive evidence that this is the case.

Conclusion

The observed geographic dispersion in life expectancy and evidence from movers between areas strongly suggest that where one lives matters for when one dies. Determining whether place health effects are large or trivially small, however, has not been accomplished until very recently. New evidence comparing movers to other movers to estimate place health effects make it reasonable to conclude that, at least for some groups, place of residence has a sizable effect on health. However, more research is needed to build on these findings and, in particular, to understand the effect of place at younger ages on long-term longevity. Although there are many plausible mechanisms through which these place effects may materialize, the question of what it is exactly that causes some places to be better for health than others has so far not been answered directly by any existing study. Given the conceptual need to have local characteristics be as good as randomly assigned, studies that use quasi-experimental regional variation are necessary to make progress on this dimension.

Can public policy take advantage of place health effects in a way that would improve public health? One possible conclusion to draw from the emergent literature is that helping individuals relocate to places that are better for health could be welfare-improving. An advantage of such a policy is that one only needs to know which places are conducive to good health rather than understand the exact mechanisms behind place health effects. However, given the observed reluctance of individuals to move to higher-wage areas (Kennan and Walker 2011), a program that offers subsidies to those who relocate to more favorable locations is likely to be very costly.

It is also unclear whether individuals who are most likely to move as a result of any given policy are those who would benefit most from positive place effects. If individuals are not taking advantage of place health effects due to imperfect information, however, a welfare improvement at a fairly low cost may be possible. If social or family ties are what bind individuals to a particular location, then any program that aims to relocate individuals to healthier places would need to be designed in a way that coordinates relocation (or perhaps improves communication and travel) of related individuals or of social networks. Viewing geographic differences in health outcomes through the lens of Roback's (1982) spatial sorting model also offers a reminder that ending up in "unhealthy" places will be at least partially the result of choices that include an array of observed and unobserved factors. Any policy that attempts to influence relocation without understanding why individuals have not already relocated themselves to places that would benefit their health runs the risk of reducing overall welfare.

An alternative policy goal would be to target health-improving policies to areas that have been shown to be detrimental to health. Without greater understanding of the mechanisms behind place health effects, however, it is unclear which local characteristics such a policy should try to improve. Additionally, given scarce social resources, it is worth considering whether policies that target some other existing inequality (perhaps in wealth or income) would be superior to policies that target life expectancy more directly. Indeed, it may be that targeting income or wealth inequality would reduce inequalities not only in life expectancy but also in other non-health dimensions.

■ *We thank the editors of JEP for detailed feedback that greatly improved the manuscript, Celestina Edleman and Yifan Wang for excellent research assistance, and Nicolai Kuminoff and Jonathan Skinner for helpful feedback. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under award numbers R21AG050795, P01AG005842, and R01AG053350. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.*

References

Abaluck, Jason, Mauricio M. Caceres Bravo, Peter Hull, and Amanda Starc. 2020. "Mortality Effects and Choice across Private Health Insurance Plans." NBER Working Paper 27578.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *Quarterly Journal of Economics* 134 (4): 1793–1844.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98–108.

Arias, Elizabeth, Loraine A. Escobedo, Jocelyn Kennedy, Chunxia Fu, and Jodi A. Cisewski. 2018. *US Small-Area Life Expectancy Estimates Project: Methodology and Results Summary*. Hyattsville, MD: National Center for Health Statistics.

Barreca, Alan I., Matthew Neidell, and Nicholas J. Sanders. 2017. "Long-Run Pollution Exposure and Adult Mortality: Evidence from the Acid Rain Program." NBER Working Paper 23524.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. 2016. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy* 124 (1): 105–59.

Bastian, Brigham, Betzaida Tejada-Vera, Elizabeth Arias, et al. 2020. "Mortality Trends in the United States, 1900–2018." National Center for Health Statistics.

Baum, Aaron, Juan Wisnivesky, Sanjay Basu, Albert L. Siu, and Mark D. Schwartz. 2020. "Association of Geographic Differences in Prevalence of Uncontrolled Chronic Conditions with Changes in Individuals' Likelihood of Uncontrolled Chronic Conditions." *JAMA* 324 (14): 1429–38.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review* 105 (2): 477–503.

Buckles, Kasey, Andreas Hagemann, Ofer Malamud, Melinda Morrill, and Abigail Wozniak. 2016. "The Effect of College Education on Mortality." *Journal of Health Economics* 50: 99–114.

Card, David, and Laura Giuliano. 2013. "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends." *Review of Economics and Statistics* 95 (4): 1130–49.

Chay, Kenneth, and Michael Greenstone. 2003a. "Air Quality, Infant Mortality, and the Clean Air Act of 1970." NBER Working Paper 10053.

Chay, Kenneth, and Michael Greenstone. 2003b. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118 (3): 1121–67.

Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li. 2013. "Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy." *Proceedings of the National Academy of Sciences* 110 (32): 12936–41.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133 (3): 1107–62.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association between Income and Life Expectancy in the United States, 2001–2014." *JAMA* 315 (16): 1750–66.

Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review* 108 (10): 3028–56.

Currie, Janet, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania. 2010. "The Effect of Fast Food Restaurants on Obesity and Weight Gain." *American Economic Journal: Economic Policy* 2 (3): 32–63.

Currie, Janet, and Matthew Neidell. 2005. "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *Quarterly Journal of Economics* 120 (3): 1003–30.

Currie, Janet, and Hannes Schwandt. 2016. "Mortality Inequality: The Good News from a County-Level Approach." *Journal of Economic Perspectives* 30 (2): 29–52.

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3 (1): 65–90.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." *American Economic Review* 109 (12): 4178–4219.

Deryugina, Tatyana, and David Molitor. 2020. "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina." *American Economic Review* 110 (11): 3602–33.

Deryugina, Tatyana, and David Molitor. 2021. "Data and Code for: The Causal Effects of Place on Health and Longevity." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor].

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–85.

Deschênes, Olivier, Michael Greenstone, and Joseph S. Shapiro. 2017. "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program." *American Economic Review* 107 (10): 2958–89.

Doyle, Joseph J. 2011. "Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far from Home." *American Economic Journal: Applied Economics* 3 (3): 221–43.

Doyle, Joseph J., John A. Graves, and Jonathan Gruber. 2017. "Uncovering Waste in US Healthcare: Evidence from Ambulance Referral Patterns." *Journal of Health Economics* 54: 25–39.

Doyle, Joseph J., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy* 123 (1): 170–214.

Dwyer-Lindgren, Laura, Amelia Bertozzi-Villa, Rebecca W. Stubbs, Chloe Morozoff, Johan P. Mackenbach, Frank J. van Lenthe, Ali H. Mokdad, and Christopher J.L. Murray. 2017. "Inequalities in Life Expectancy among US Counties, 1980 to 2014: Temporal Trends and Key Drivers." *JAMA Internal Medicine* 177 (7): 1003–11.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou. 2017. "New Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy." *Proceedings of the National Academy of Sciences of the United States of America* 114 (39): 10384–89.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review* 111 (8): 2697–2735.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127 (3): 1057–1106.

Fletcher, Jason M. 2010. "Social Interactions and Smoking: Evidence Using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects." *Health Economics* 19 (4): 466–84.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields. 2005. "The Causal Effect of Income on Health: Evidence from German Reunification." *Journal of Health Economics* 24 (5): 997–1017.

Fuchs, Victor R. 1974. *Who Shall Live? Health, Economics, and Social Choice.* New York: Basic Books.

Galama, Titus J., Adriana Lleras-Muney, and Hans van Kippersluis. 2018. "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence." NBER Working Paper 24225.

Gibson, John, Steven Stillman, David McKenzie, and Halahingano Rohorua. 2013. "Natural Experiment Evidence on the Effect of Migration on Blood Pressure and Hypertension." *Health Economics* 22 (6): 655–72.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.

Heutel, Garth, Nolan H. Miller, and David Molitor. 2020. "Adaptation and the Mortality Effects of Temperature across US Climate Regions." NBER Working Paper 23271.

Hut, Stefan. 2020. "Determinants of Dietary Choice in the US: Evidence from Consumer Migration." *Journal of Health Economics* 72.

Johnson, Janna E., and Evan J. Taylor. 2019. "The Long-Run Health Consequences of Rural–Urban Migration." *Quantitative Economics* 10 (2): 565–606.

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1): 211–51.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. 2016. "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health." *Review of Economics and Statistics* 98 (2): 350–66.

Lindahl, Mikael. 2005. "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income." *Journal of Human Resources* 40 (1): 144–68.

Macintyre, Sally, Anne Ellaway, and Steven Cummins. 2002. "Place Effects on Health: How Can We Conceptualise, Operationalise and Measure Them?" *Social Science & Medicine* 55 (1): 125–39.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.

Miller, Sarah, Norman Johnson, and Laura R. Wherry. 2019. "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data." NBER Working Paper 26081.

Mokdad, Ali H., Katherine Ballestros, Michelle Echko, Scott Glenn, Helen E. Olsen, Erin Mullany, Alex Lee et al. 2018. "The State of US Health, 1990–2016: Burden of Diseases, Injuries, and Risk Factors among US States." *JAMA* 319 (14): 1444–72.

Molitor, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy* 10 (1): 326–56.

Montez, Jennifer Karas, Jason Beckfield, Julene Kemp Cooney, Jacob M. Grumbach, Mark D. Hayward, Huseyin Zeyd Koytak, Steven H. Woolf, and Anna Zajacova. 2020. "US State Policies, Politics, and Life Expectancy." *Milbank Quarterly* 98 (3): 668–99.

Montez, Jennifer Karas, Anna Zajacova, Mark D. Hayward, Steven H. Woolf, Derek Chapman, and Jason Beckfield. 2019. "Educational Disparities in Adult Mortality across US States: How Do They Differ, and Have They Changed Since the Mid-1980s?" *Demography* 56 (2): 621–44.

Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *Journal of Political Economy* 114 (5): 871–904.

Murray, Christopher J.L., Sandeep C. Kulkarni, Catherine Michaud, Niels Tomijima, Maria T. Bulzacchelli, Terrell J. Iandiorio, and Majid Ezzati. 2006. "Eight Americas: Investigating Mortality Disparities across Races, Counties, and Race-Counties in the United States." *PLoS Medicine* 3 (9).

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. 2017. "The Gift of Moving: Intergenerational Consequences of a Mobility Shock." NBER Working Paper 22392.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics* 116 (2): 681–704.

Schwandt, Hannes. 2018. "Wealth Shocks and Health Outcomes: Evidence from Stock Market Fluctuations." *American Economic Journal: Applied Economics* 10 (4): 349–77.

Schwandt, Hannes, and Till M. von Wachter. 2020. "Socioeconomic Decline and Death: Midlife Impacts of Graduating in a Recession." NBER Working Paper 26638.

Skinner, Jonathan, and Elliott Fisher. 2010. "Reflections on Geographic Variations in US Health Care." https://data.dartmouthatlas.org/downloads/press/Skinner_Fisher_DA_05_10.pdf.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E. Wennberg, and Elliott S. Fisher.

2010. "Regional Variations in Diagnostic Practices." *New England Journal of Medicine* 363 (1): 45–53.
Sullivan, Daniel, and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data." *Quarterly Journal of Economics* 124 (3): 1265–1306.
Woolf, Steven H, and Heidi Schoomaker. 2019. "Life Expectancy and Mortality Rates in the United States, 1959–2017." *JAMA* 322 (20): 1996–2016.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 171–196

When Innovation Goes Wrong: Technological Regress and the Opioid Epidemic

David M. Cutler and Edward L. Glaeser

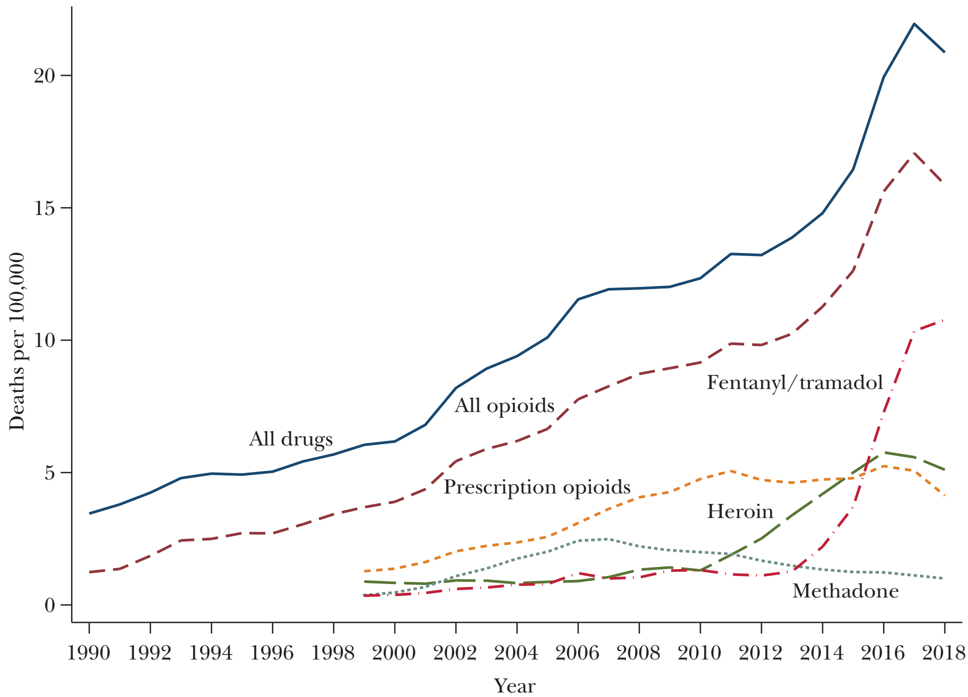
The fourfold increase in the death rate from opioid drugs between 2000 and 2017, illustrated in Figure 1, is an American health crisis rivalling even the COVID-19 pandemic. Nearly 500,000 people died from opioid overdoses between 1999 and 2019, and in 2019, more people died from opioids (49,860) than from motor vehicle accidents (38,800) or breast cancer (42,281) (Centers for Disease Control and Prevention, National Center for Health Statistics 1999–2019). The increase in drug overdose deaths is a major reason for recent declines in US life expectancy (Kochanek, Arias, and Bastian 2016; Woolf and Schoomaker 2019) and has contributed to the longer-term increase in mid-life mortality among white non-Hispanics first emphasized by Case and Deaton (2015). The opioid crisis has also exacerbated the link between lifespan and education; the opioid death rates are far higher for those without a Bachelor of Arts than for those with one (Ho 2017). Opioid deaths rose during the COVID pandemic, despite the sharp reductions in mobility (Goodnough 2021).

At its heart, the opioid story is one of technological regress. It was hoped that a new wave of opioid-based drugs could end America's longstanding scourge of untreated pain, just as antihypertensives, cholesterol-lowering agents, and antidepressants brought therapy to millions of previously untreated people with high blood pressure, high cholesterol, and mental illness. It was not to be.

■ *David M. Cutler is Otto Eckstein Professor of Applied Economics and Edward L. Glaeser is Fred and Eleanor Glimp Professor of Economics, both in the Economics Department at Harvard University, Cambridge, Massachusetts. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are dcutler@harvard.edu and eglaeser@harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.171.

Figure 1
Trends in Age- and Sex-Adjusted Drug Deaths and Opioid Deaths, 1990–2018



Note: Data are from the Centers for Disease Control and Prevention, National Vital Statistics System (1990–2018). Drug deaths after 1999 were identified based on *International Classification of Diseases, 10th edition* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Overdoses by category were identified by multiple-cause-of-death codes T40.1 (heroin), T40.2 (prescription opioids = natural and semisynthetic opioids), T40.3 (methadone), and T40.4 (fentanyl/tramadol = synthetic opioids other than methadone) (Hedegaard, Miniño, and Warner 2018). Total opioid deaths also included code T40.6 (other/unspecified narcotics). Drug deaths before 1999 were identified based on *International Classification of Diseases, 9th edition* underlying cause-of-death codes E850-E858, E950.0-E950.5, E9620, and E980.0-E980.5. Opioid deaths before 1999 were identified from underlying cause-of-death-codes E850.1-E850.2 and 305.5, as well as multiple-cause-of-death codes 965.00-965.09 (Fingerhut and Cox 1998). Deaths involving more than one opioid category are counted in both. To account for the change from ICD-9 codes (1990–1998) to ICD-10 codes (1999–2017), the following comparability ratios were applied to ICD-9 codes E850-E858, E950-E950.5, E9620, and E980.0-E980.5 (respectively) in the calculation of total drug deaths: 1.0365, 1.0013, 0.9870, and 1.0417 (Miniño et al. 2006). Total opioid deaths were adjusted upward by about 20 percent (comparability ratio = 1.195) (Hoyert et al. 2001). Deaths were age- and sex- adjusted to the US 2000 population, using population data from the Surveillance, Epidemiology, and End Results (SEER) Program (NIH 2021). Deaths from specific causes were adjusted for underreporting following Ruhm (2018), as described in the online Appendix avaialble with this paper at the *JEP* website.

The opioid epidemic began with the availability of OxyContin in 1996. OxyContin was portrayed as a revolutionary wonder drug: because the painkiller was released only slowly into the body, relief would last longer and the potential for addiction would decline. From 1996 to 2011, legal opioid shipments rose six-fold. But the hoped-for benefits proved a mirage. Pain came back sooner and stronger than expected. Tolerance built up, which led to more and higher doses. Opioid use led to opioid abuse, and some took to crushing the pills and ingesting the medication all at once. A significant black market for opioids was born.

Fifteen years after the opioid era began, restrictions on their use began to bind. From 2011 on, opioid prescriptions fell by one-third. Unfortunately, addiction is easier to start than stop. With reduced access to legal opioids, people turned to illegal ones, first heroin and then fentanyl, which has played a dominant role in the recent spike in opioid deaths.

In this essay, we begin with a brief sketch of the history of opioids and the rise of their use in the United States since about 1995. Our main focus is on the positive question of how demand for and supply of opioids produced the epidemic. In considering demand, we look at available measures of physical and mental pain, despair, and the opportunity cost of time, which is associated with joblessness and social isolation. We show that changes in demand-side factors alone, such as physical pain, depression, despair, and social isolation can only explain a small fraction of the increase in opioid use and deaths from 1996 to 2012. However, we also find that patterns of demand helped to shape the locations in which the opioid crisis became most severe.

The dominant changes in opioid supply started with modest technological and marketing innovations in the legal sector, which was followed by a burst of entrepreneurship in the illegal sector. In the legal market, physicians who cared about treating the impaired were persuaded by a time-release system and a highly effective marketing campaign that the new opioids were truly safer than the older ones, and they started prescribing. While the opioid crisis did not begin with supply shifts in the illegal market, technological and institutional changes within that market furthered the epidemic. The introduction of fentanyl and the rise of Asian fentanyl exports appears to be a narcotic variant of the broader China trade shock that occurred in the early 2000s, in which declining transport costs and East Asian industrial expertise flooded American markets and displaced the opium producers of Mexico (Grandmaison, Morris, and Smith 2019).

Opioid prescriptions are now down substantially from their peak. However, even if the reduction in legal opioid prescriptions since about 2011 reduces the flow of new addicts, the stock of existing addicts will continue to seek supply, even when it means substituting more dangerous illegal sources. Thus, the opioid epidemic is likely to be with us for some time to come.

The Opioid Crisis and Its History

Opiates are naturally occurring substances that can be extracted from the opium poppy plant. The plant can be smoked directly or purified into more potent opiates including legal drugs like codeine and morphine or illegal drugs like heroin. The term *opioid* refers also to semi-synthetic drugs such as oxycodone (the key ingredient in OxyContin), hydrocodone, and hydromorphone, and to fully synthetic drugs such as fentanyl and methadone.

The various derivatives of opium are chemically similar, but their potency varies greatly (Centers for Medicare and Medicaid Services 2017). Codeine has 15 percent the potency of morphine, while oxycodone has 150 percent of morphine's potency. Heroin is roughly three times as potent as morphine, and fentanyl is 100 to 200 times more potent than morphine, with the variation arising because the potency of illegal drugs varies from batch to batch.

Opioids relieve pain and make people feel calm and happy. They also depress basic bodily functions such as respiration and cardiac activity. For this reason, a dose that is far enough above the typical amount can lead to death, even among tolerant users. Unfortunately, the line between euphoria and death is not very wide by therapeutic standards. Gable (2004) finds that for intravenous heroin, the lethal dose is only six times the effective dose, making it the most dangerous of common drugs.

The Ultimate Addictive Good

Opium has been used to stimulate pleasure and relieve pain since at least 3400 BCE. There are Sumerian references to the "joy plant" (Booth 1996; Saunders 2014). Opium was well-known to civilizations from Greece to Egypt to Persia to India, both for its beneficial effect and possible overdoses. Hippocrates (460–377 BCE), the father of modern medicine—from whom we have the Hippocratic Oath—frequently mentioned the poppy in his remedies. Herakleides of Pontus (~340 BCE) in his book *On Government* describes how one island's inhabitants regularly committed suicide "by means of the poppy" (Saunders 2014). Avicenna's 11th century *Canon of Medicine* warned that "the most powerful of the stupefacients is opium," which made it a useful painkiller, but that it was also "definitely poisonous" (Avicenna 1998).

Opium became a major trade good and a source of conflict during the Age of Exploration. Britain and China fought two 19th-century "opium wars," which ended with the British protecting their right to sell opium in China. Perhaps one in five Chinese men were opium users early in the 20th century (United Nations Office on Drugs and Crime 2010), and the Chinese opium epidemic only ended when the Communists imposed draconian restrictions on consumption.

Opium's combination of danger and pleasure has led to repeated cycles of innovation, addiction, and correction, which begin when entrepreneurs produce an allegedly safer opioid. However, when purchasers begin consuming the new drug, they discover that this new innovation is as addictive and deadly as the old forms of opium. New consumers avoid the drug or are prohibited from using it. Existing users pass away, and the fad dies down—until memories fade and the cycle begins again. London physician Thomas Sydenham combined opium with alcohol in 1676 to produce laudanum, a wonder drug that eliminated almost all forms of pain—and which became its own substance abuse problem. Twenty-five years later, in 1701, Dr. John Jones wrote in "The Mysteries of Opium Reveal'd" that long term use of opium generates an "inability or listlessness to do any things except it be while the Opium operates," but that quitting opium use could leave to "intolerable . . . anxieties," and even a "miserable death."

In 1804, Friedrich Serturner separated "morphine" from opium. He believed that he had discovered a safe medication, but he would himself become an addict. Several decades later, Merck produced the drug commercially. Morphine and opium were widely used as painkillers during the US Civil War, and morphine addiction was termed the "Soldier's Disease" in the last third of the 19th century.

By 1872, the Annual Report of the State Board of Health of Massachusetts noted that "the sulphate of morphia seems to be growing in favor," and that "this salt is not only taken internally, but is sometimes used hypodermically." The Report repeats the canard that morphine is "free from the more objectionable properties of opium," but also reports (p. 167) a comment from a state assayer that "among the most dangerous preparations of morphia are those now prescribed and sold by uneducated or villainous individuals as so-called 'cures' for persons afflicted with the uncontrollable appetite for opium."

Pierre Robiquet isolated codeine in 1832, and it remains the most commonly prescribed opiate today. Felix Hoffmann at Bayer was trying to produce codeine when he stumbled upon heroin, a more potent form of morphine. The Bayer company marketed heroin, claiming: "Heroin is completely devoid of the unpleasant and toxic effects of opium derivatives." The *Boston Medical and Surgical Journal* (the forerunner of the *New England Journal of Medicine*) informed its readers in 1900 that heroin "possesses many advantages over morphine as a respiratory sedative," especially an "absence of danger of acquiring the habit" (Daly 1900, p. 190). Heroin was sold to suppress coughs, relieve the burden of childbirth and war injuries, prepare for anesthesia, and control certain mental disorders. As the dangers of heroin became clear, Bayer phased out its promotion in favor of another new compound, acetylsalicylic acid (aka "aspirin"), also synthesized by Felix Hoffmann.

Semi-synthetic and synthetic opioids were developed in the 20th century. In 1916, two German scientists produced oxycodone, and it became a popular painkiller for the Nazis; Hitler and Goering both appear to have been enthusiasts (Ohler 2017). Fully synthetic opioids came later still. Fentanyl was created in Belgium in 1959, and Tramadol was developed in Germany in 1962.

Opioids are an extreme example of the addictive goods analyzed by Becker and Murphy (1988). They have strong intertemporal complementarity in consumption: past use greatly increases the marginal benefit of current use. Further, there is a large tradeoff between short-term mood advantages and longer-term downsides. The longer-term costs from opioids are not direct health costs like the lung damage generated by cigarettes on smokers, but are indirect and mediated by addiction or "tolerance." When addicts attempt to satisfy their habit, especially by using illegal opioids, they pay financial costs and face risk of overdose and death. Nutt et al. (2007) surveyed experts to determine a scale of harm for 20 different drugs and found that heroin generated the highest level of dependence and risk of an overdose.[1]

The Illegal Supply of Opioids before the Epidemic

US legislation in the early 20th century sought to limit opioid use as a way to reduce widespread morphine and heroin abuse. The Harrison Narcotics Tax

[1] Of the world's most widely used drugs, only caffeine appears to have practically no well-documented long-term health costs.

Act of 1914 required consumers to have a doctor's prescription to gain access to opiates, and doctors were typically skeptical about prescribing drugs that appeared to be addictive. After 1917, the Act was interpreted to mean that doctors should not prescribe opiates to addicts merely to maintain their habits. Heroin was banned in 1924, owing to its alleged impact on crime. Codeine remained a mainstay of cough suppressants, but medical prescriptions of other opiates plummeted.

Legal restrictions did not end the supply of opioids; rather, that supply moved underground. Before World War II, at least in Chicago, opium smoking was far more common than heroin injection and it was concentrated in the city's Chinatown (Dai 1937). Between the 1950s and 1972, America's heroin supply appears to have been primarily produced from poppies grown in Turkey and smuggled in through the "French Connection." That route was disrupted by law enforcement in 1972, and the price of heroin correspondingly rose (Brown and Silverman 1974). The combination of high prices and the shutdown of the European supply chain attracted other entrepreneurs, from Southeast Asia's Golden Triangle to Afghanistan's Golden Crescent, and later, Mexico.

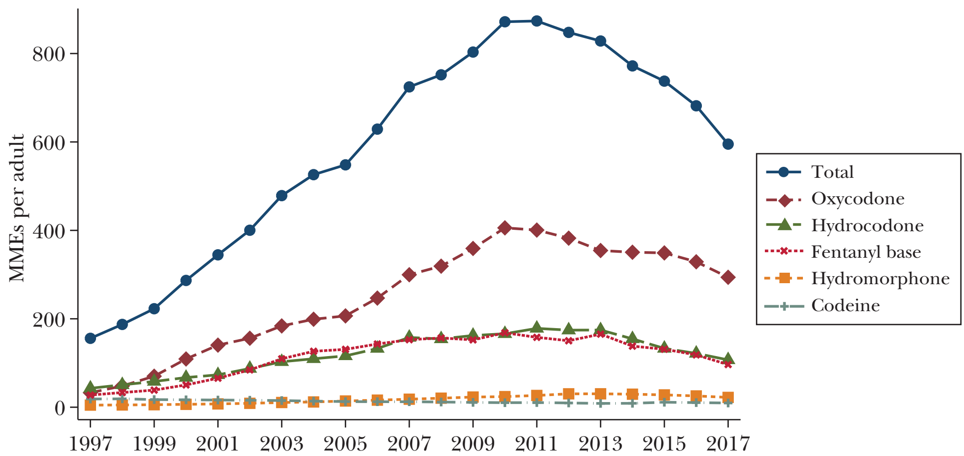
The long-run supply of heroin seems to be quite elastic. Heroin prices fell over 80 percent in real terms between 1981 to 2001. However, the new supply sources boosted consumption, and heroin-related deaths started rising again in the early 1990s. Still, none of these heroin crises saw death rates anywhere near those that appeared after 2010.

The Rise of Opioid Use

Trends in legal opioid use in the past two decades are shown in Figure 2, which presents aggregate shipments of opioids per adult—in milligrams of morphine equivalents, or MMEs—from 1997 through 2017. The data underlying Figure 2, and all data on legal opioid shipments that we present, are from the Drug Enforcement Administration's Automation of Reports and Consolidated Order System (ARCOS) (1999–2017), which details shipments by product and three-digit zip code. We aggregate the ARCOS data to counties or the nation, using a zip-to-county crosswalk from the US Census Bureau (US Census Bureau 2010) (for details, see the online Appendix). Fifty MME is a typical daily dose for a person in pain. Thus, total opioid supply in 1997 was roughly three days of typical use per adult. By 2011, the supply was 20 days per person—roughly one prescription per adult per year. The overall growth was 461 percent. OxyContin was a major part of the total. Oxycodone shipments rose 27 percent annually from 1997 to 2011. Shipments of other opioids rose as well, though none to quite the same extent.

The increased use of opioids involved changes on the extensive and intensive margin (again, see the online Appendix for details). Our primary source of prescription data is the Medical Expenditure Panel Study (MEPS), an ongoing survey of the non-institutional population since 1996 (Agency for Healthcare Research and Quality 2021). MEPS data show that opioid prescriptions per capita doubled between 1996 and 2010. Data from IMS Health show similar trends

Figure 2
Trends in Opioid Shipments per Adult, 1997–2017



Note: The figure shows milligrams of morphine equivalents per adult in the United States from 1997 to 2017. Data are from the DEA ARCOS database. The total omits methadone and buprenorphine, which are often used in treating opioid abuse, along with some smaller opioids for which data are not available in all years. Data for hydromorphone, codeine, morphine, and fentanyl base was missing in 2000 and was imputed using data in 1999 and 2001, assuming a constant growth rate. The same was done for total shipments.

(Volkow 2014). One-quarter of the increase in prescriptions came from the extensive margin of more people being prescribed medications; three-quarters was due to the intensive margin of more scripts per person. In addition to more prescriptions, data from the Massachusetts Prescription Drug Monitoring Program (PDMP) show that days supplied per prescription increased by a quarter from 2000 to 2012 (Massachusetts Department of Public Health 2000–2018).

The increase in the number of people receiving prescriptions, the number of prescriptions per recipient, and the number of pills per script does not add to the total increase in opioid shipments. It is possible that the gap is due to high volume "pill mills," not all of which might be recorded in national survey data.

As one metric for comparison, we focus on the share of people with two or more prescriptions in the ~2½ years people are asked about in the survey to capture heavy use.[2] There are stark differences in heavy use by education. In 2009, 13 percent of people without a college degree had more than one opioid script, compared to 9 percent of people with a college degree. Opioid use also began to fall earlier for people with college degrees. In 1998, opioid use was the same in rural and urban areas, but then it rose more rapidly in rural areas. By 2012, 15 percent of rural residents had more than one opioid script, as opposed to 10 percent of urbanites. Heavy use declined more rapidly in urban areas after 2012, further increasing

[2] The online Appendix shows variability in opioid use along several dimensions, using the MEPS data.

the gap between country and city. Differences by labor force participation are also marked: 20 percent of people aged 25–44 who were out of the labor force were heavy opioid users in 2012, compared to 10 percent for blue-collar workers and 6 percent for white-collar workers.

The Medical Expenditure Panel Study data also allows us to see whether opioids served as a substitute for other medications. To examine this, we take advantage of the panel nature of the data. People are in MEPS for five rounds, each lasting about half a year. Starting with the 2001 panel, people in the second round were asked: "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?" We count people as in pain if they answer "moderately," "quite a bit," or "extremely," as opposed to "not at all" or "a little bit."

We sample people who are in pain in the second round of the survey, but who were not taking opioids, antidepressants, or anxiolytics in the first round. Among that group, the share of opioid users increased dramatically from about 6 percent in 2001 to about 9 percent in 2009. The use of antidepressants and anxiolytics (anti-anxiety medications) was generally flat. Use of other pain medications, for example high dose prescription non-steroidal anti-inflammatories, fell. Thus, is appears that opioids primarily substituted for less powerful pain relievers but not for antidepressants or anti-anxiety medications.

Shifting Demand and the Rise of Opioids

Opioid use is determined by the interaction of demand and supply in the linked markets for legal and illegal drugs. The observed increase in opioid use can reflect an increase in demand, in supply, or both. Demand factors can explain the course of the opioid epidemic either because rising opioid use reflect shifts in demand or because rising supply drives up national opioid use, but local demand factors determine where that supply has the most impact. Here we address whether demand shifts on their own can explain the national rise in opioid use. We consider four potential demand shifters for opioids: physical pain, depression, despair, and social isolation.

Demand-Side Forces: Physical Pain, Depression, Despair, and Opportunity Cost of Time

Jeremy Bentham (1789, p. i) famously noted: "Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*." Opioid use replaces pain with pleasure. As pain increases, one would naturally expect opioid use to increase as well, as happened after the US Civil War.

Opioid use might also appeal to individuals who are in psychic pain: Case and Deaton (2017) famously termed America's rising middle-aged mortality rates "deaths of despair." However, pain caused by physical impairment differs in a central way from pain caused by mental impairment. Physical pain, suffered on the factory floor or the battlefield, may be unrelated to any other personal attribute. Mental pain, generated by disorders such as depression or a consequence of social and economic changes, often coexists with other characteristics that may either increase or decrease the demand for opioids. Anxiety, for example, might reduce opioid use by generating increased fear of addiction or an inability to take actions, like finding a cooperative doctor or dealer.

Opioids may relieve pain and lessen despair, but for most consumers, use comes at a cost of diminished attention and energy. Serious opioid consumption produces lethargy and diminishes interest in other people. Consequently, the demand for opioids should be higher among people with a lower opportunity cost of time. An extreme version of this hypothesis is that opioids can be a complement to doing nothing, because the short-term pleasure generated by opioids is large and independent of most other activities. For much of the past decade, more than 15 percent of 25–54 year-old men have not been employed. The employment-to-population ratio for these "prime-aged" men has been less than 75 percent in many parts of America's eastern heartland, where opioid use has been severe (Austin, Glaeser, and Summers 2018). Krueger (2017) reports that one-half of jobless men take some form of painkillers.

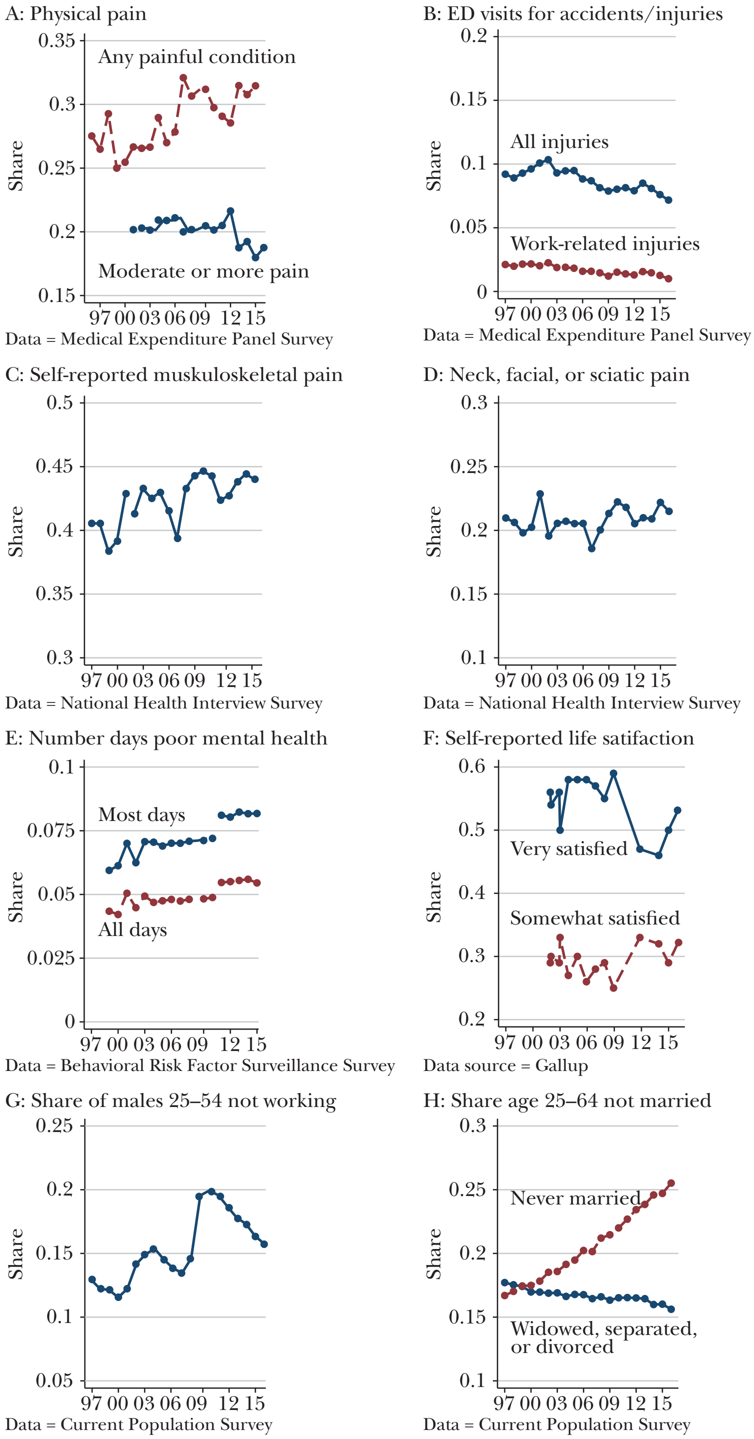
Patterns of Demand-Side Variables

Figure 3 shows national-level trends in four variables related to these explanations: physical pain, depression, despair, and social isolation. Trends in these variables are age- and sex-adjusted to the 2000 US population when possible. Panel A shows two measures of pain in the Medical Expenditure Panel Study data. The first is the measure of pain interfering with work in the past four weeks, noted above. The second is whether the person reports one of eleven painful conditions, like arthritis or back pain. About 20 percent of people report that pain interfered with their normal work at least moderately in the past four weeks and 25 percent report a painful condition. The share of people reporting that pain interfered with their work is relatively constant, but the prevalence of reporting at least one painful condition increased 14 percent. The difference between these two may imply that pain treatment is somewhat effective, or it could be that pain is now given a diagnosis where formerly it was not. Panels B, C, and D report the share of adults visiting the emergency department for any injury or a workplace injury; the share of adults with joint pain, back pain, or neck pain; and the share of adults with neck, facial, or sciatic pain, the latter two taken from the National Health Interview Survey (Blewett et al. 2019). As in MEPS, about one-third report musculoskeletal pain, and this increased about 12 percent over the time period. Neck, facial, or sciatic pain rose about 5 percent. Emergency room visits for injuries declined.

Panel E shows mental health impairment, the share of people with poor mental health all or most days from the Behavioral Risk Factor Surveillance System Annual Survey (CDC 2021a). Trends are split in 2011, where there was a change to BRFSS weighting methodology, and as is described in the online Appendix.[3] Each has

[3] See "Can the 2010 BRFSS dataset be compared with 2011 dataset?" from the Centers for Disease Control and Prevention website at https://www.cdc.gov/surveillancepractice/reports/brfss/brfss_faqs.html.

Figure 3
Trends in Pain, Mental Health, Despair, and Social Isolation



Note: For sources, see text. Data are adjusted to the 2000 population by age and sex except for longer-term data from Gallup in Panel F.

increased over time, by 41 percent and 85 percent respectively. Panel F shows a measure of despair, average life satisfaction (Gallup 1997–2015). The survey shows that 86 percent of people were satisfied with their life in 1996 and 2019; the share has remained between 73 and 87 percent since 1979. Life satisfaction rebounded after the Great Recession.

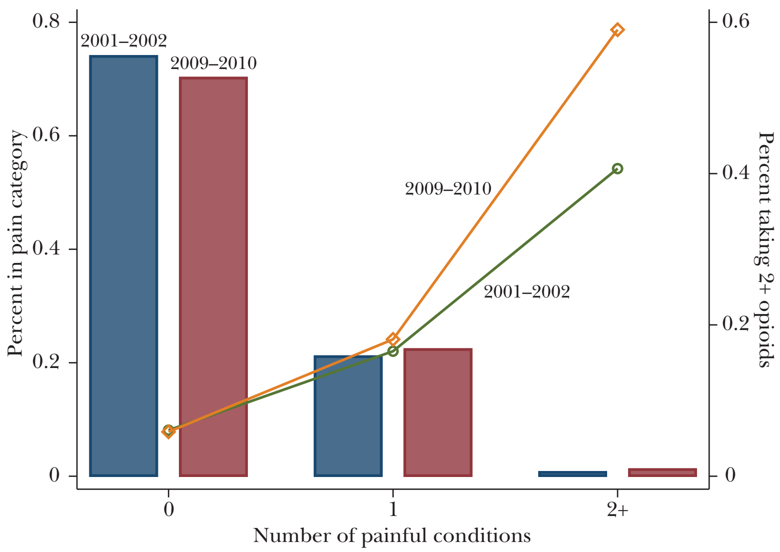
The last two panels (G and H) capture social isolation and the opportunity cost of time: the share of the 25–54-year-old male population that is not employed and the share of the population aged 25–64 that is never married, drawn from the Current Population Survey (Flood et al. 2020). Both of these measures have increased, the latter especially so. The percent of 18–34-year-olds who are married fell from 58 percent in 1978 to 29 percent in 2018. The share of Americans who live alone increased from 13 percent in 1960 to 28 percent in 2016. Putnam (1995) documents the widespread decline of many different forms of social connection, from bowling leagues to fraternal clubs. Quinones (2015) explicitly links the rise of opioid abuse to weakening social ties, particularly in the eastern heartland.

Demand and Opioid Initiation

In Figure 3, many of the variables increase in prevalence over time, but they do not increase enough to offer a promising source of explanation for the four-fold increase in opioid deaths or the six-fold increase in opioid shipments from 1996 to 2012. Figure 4 shows this informally, focusing on the 2001–2002 and 2009–2010 cohorts. The first cohort is the earliest with good pain data; the latter is at the height of opioid prescribing. The figure shows that the share of people with 2+ painful conditions rises across the two time periods, but the share of people with 2+ conditions who have at least two opioids scripts in the 2½ years they are in the Medical Expenditure Panel Study rises by even more. The lion's share of the growth in opioid use comes from increased number of prescriptions, holding the number of pain categories fixed, not from the increase in the number of people reporting multiple forms of pain.

As a more formal test, we sample people in the Medical Expenditure Panel Study who are opioid-naïve in the second round. For this group, we record several measures of potential demand for opioids: whether the person reports pain, based on the two measures above; how much of the past four weeks the person felt downhearted and depressed, which we scale from zero to one; and several measures of social isolation, including labor force participation and marital status. We relate each of these variables to the onset of heavy use opioids (two or more opioid prescriptions) after the second round. The regressions are available in the online Appendix. All of these variables are associated with subsequent use of opioids. Controlling for basic demographics, health care access, and other health behaviors, an individual in extreme pain has a 15 percentage point higher probability of initiating multiple opioid prescriptions than an individual with no pain, a person who reported being downhearted and depressed all of the time in the past four weeks is 5 percentage points more likely to initiate heavy opioid use than one reporting no time spent downhearted and depressed, and a person who is widowed, separated, or divorced is 2 percentage points more likely to initiate heavy opioid use than a married person.

Figure 4
Pain and the Initiation of Opioids



Note: Data are from the Medical Expenditure Panel Study. Painful conditions include sickle cell anemia, headache (including migraine), nonspecific chest pain, rheumatoid arthritis and related disease, osteoarthritis, other nontraumatic joint disorders, spondylosis (including intervertebral disc disorders other back problems), joint disorders and dislocations (trauma-related), all fractures, sprains and strains, and abdominal pain.

However, these variables did not increase enough over time to explain the growth in opioid use, implying that they cannot explain the growth in opioid use. Regressions including the report of painful conditions explain 20 percent of the growth in heavy opioid use; regressions using pain interfering with work explain only 4 percent of that increase. The same pattern appears when looking at explaining the growth in the use of any opioid.

Because the Medical Expenditure Panel Study does not have a good measure of despair, we turn to the Midlife in the US Survey (MIDUS) to examine the impact of despair on opioid use (Brim et al. 1995–1996; Ryff et al. 2004–2006; Ryff et al. 2013–2014). MIDUS is a panel of roughly 7,000 people who were interviewed in 1995–1997, in 2004–2005, and a third time in 2013–2015. We consider whether an opioid-naïve person in one survey round (that is, round 1 or 2) becomes a user of prescription pain relievers by the next round, controlling for basic demographics. MIDUS asks about prescription pain relievers in general, not opioids in particular. The bulk of prescription pain relievers, but not all, are likely to be opioids. On average, 27 percent of people in waves 2 and 3 take prescription pain relievers. MIDUS asks a number of questions on health and outlook. We combine groups of questions into summary statistics meant to capture the categories of pain, negative affect, despair, economic insecurity, and social isolation. Measures that feed into despair include life satisfaction, social integration, and perceived contributions to society. Economic insecurity includes measures of their financial situation and difficulty paying bills.

The results based on the Midlife in the US Survey data for pain and negative affect are similar to those in the Medical Expenditure Panel Study. People in pain are more likely to initiate prescription pain relievers, as are people with negative affect; the impact of pain is quantitatively larger. Despair and economic insecurity matter for opioid initiation as well. A one standard deviation increase in despair increases the probability of subsequent prescription pain reliever use by 3 percentage points; a one standard deviation increase in economic insecurity leads to a 6 percentage point increase in prescription pain reliever initiation. Taken as a whole, however, changes in pain, negative affect, despair, and economic insecurity predict only one-quarter of the increase in prescription pain reliever use. In the online Appendix, we report trends in a wider variety of health, pain, and despair metrics from MEPS, NHIS, and BRFSS. As has been shown by Case and Deaton (2020) and others (Blanchflower and Oswald 2020; Case, Deaton, and Stone 2020; Nahin et al. 2019; IOM 2011), many metrics for pain and some metrics of mental health impairment and despair have increased over time, although others have declined. The increases are larger for people without a college degree. However, even changes in the metrics with the highest growth in the past two decades, where increases have been roughly 20–50 percent, pale in comparison to the rise in opioid deaths.

Other Studies

A number of studies have examined the effect of economic change on opioid use and abuse (for a review, see Maclean et al. 2020). The general finding confirms the results above: economic change over the past few decades is related to opioid overdose deaths, but the impact of economic changes on the rise in overall opioid use is modest. For example, Pierce and Schott (2020) estimate that an increase from the 25th to 75th percentile in a county's import competition from China (due to the permanent normal trade relations bill in 2000) was associated with between 2 and 3 additional drug overdose deaths per 100,000 people, less than 20 percent of the increase in the drug overdose death rate between 1999 and 2018. Similarly, Ruhm (2019) and Case and Deaton (2017) estimate that medium term economic changes such as unemployment rates and median income, have only a minor effect on opioid deaths. Ruhm (2018) estimates that from 1999 to 2015, changes in unemployment, poverty, median household incomes, home prices, and exposure to import competition—taken together—explain fewer than 10 percent of the increase in opioid deaths.

One paper reaching a different conclusion is Charles, Hurst, and Schwartz (2019a). That paper uses state-level data to estimate that the decline in manufacturing share of employment between 2000 and 2015 could explain virtually all of the increase in opioid deaths over that time period. In the online Appendix, we explore this relationship in more detail (we are grateful to Charles, Hurst, and Schwartz for providing us their data and replication code) (Charles, Hurst, and Schwartz 2019b). Data on changes in manufacturing come from the US 2000 Census and 2014–2016 American Community Survey , which we downloaded from IPUMS USA (Ruggles et al. 2021). Because the data are at the state level, Charles, Hurst, and Schwartz do not include other controls in their regression. Estimating the model at the commuting zone level and including basic demographic controls such as population age shares and median income eliminates the relationship between manufacturing decline and opioid deaths. We thus conclude that the direct effect of economic change on opioid deaths is modest. Changes in supply seems to be far more likely causes of the opioid epidemic than changes in demand-side factors including pain and despair.

The Changing Supply of Opioids

If increases in demand do not explain the increase in opioid use, the obvious alternative explanation is supply. Indeed, the recent opioid cycle is reminiscent of the supply-driven cycles seen for morphine in the 19th century and heroin in the early 20th century. In each of these cases, a pharmaceutical company produced a new and supposedly safer version of opium. Consumers bought the new drug, only to learn that it was no less addicting. Demand falls until the stock of addicts decline and memories fade, whereupon the cycle starts anew. The history of Purdue Pharma and OxyContin after 1996 follows a similar pattern, though technology has had an extra impact pushing the current cycle into illegal use.

Creating an Epidemic

OxyContin was approved by the Food and Drug Administration in December 1995, and Purdue began marketing in earnest in 1996. Three factors enabled Purdue to turn OxyContin into a blockbuster drug. First, Purdue managed to differentiate OxyContin sufficiently from past opioids, both because the semi-synthetic opioid oxycodone had less of a history than did morphine, and through the delayed-release "Contin" system. The time release system, it was hypothesized, would moderate the amount of the opioid received at any point in time, which would reduce the risk of dependence and increase the time between needed doses.[4] At the time of FDA approval and even after, no clinical trials backed up this theory.

Second, Purdue was a dynamo at drug marketing. Arthur Sackler, the oldest of the three Sackler brothers who owned Purdue, revolutionized pharmaceutical sales. He advocated "detailing, free samples, free food and drink, flashy journal advertising and mailings" (Podolsky, Herzberg, and Greene 2019, p. 1786). The 2019 lawsuit by the Commonwealth of Massachusetts against Purdue details cases such as a doctor who was visited by Purdue representatives more than 600 times after 2008

[4] In 1984, Purdue Pharma tried this approach with MS Contin, which delivered morphine slowly into the body. The forthcoming expiration of that patent, combined with the perception by physicians that morphine was too potent to give to patients on a long-term basis, led Purdue to search for other formulations (Sarpatwari, Sinha, and Kesselheim 2017).

and was given a consulting contract to promote Purdue opioids. That doctor allegedly prescribed hundreds of thousands of Purdue opioid pills, generating nearly $1.5 million of revenue for Purdue.

Third, Purdue's sales pitch rode the wave created by a nascent medical movement focused on the alleviation of pain. In 1973, the anesthesiologist John Bonica convened a meeting of pain specialists and founded the International Association for the Study of Pain, which had its own peer-reviewed journal: *Pain*. Dame Cicely Saunders, a hospice pioneer, advocated using morphine to alleviate the suffering of dying patients: obviously, in the case of dying patients, potential long-run costs of opioid use are not relevant. Purdue and other makers of pain medications provided financial support for at least one chapter of the International Association for the Study of Pain: the American Pain Society, which shut down in 2019 under a blizzard of lawsuits claiming it was a front company for opioid manufacturers and should be partly liable for the opioid epidemic.

Traditionally, opioid pain relief was considered for patients with end-stage cancer or acute trauma. Addiction was of little consequence for the first group and use for the latter group was generally limited to inpatient care. Thus, the real shift for OxyContin was in the use of opioids for the much larger group of people with chronic pain in outpatient settings.

The key to making this switch was overcoming the fear of physicians that such patients would become addicted to opioids. This was easier where there were more cancer patients; Arteaga and Barone (2021) show that opioid shipments and opioid deaths rose more rapidly in areas where cancer rates were initially higher. To win over the doctors, Purdue promoted a 1980 letter to the *New England Journal of Medicine* claiming that among "11,882 patients who received at least one narcotic preparation, there were only four cases of reasonably well documented addiction in patients who had no history of addiction" (Porter and Jick 1980). This finding was among inpatients but was taken to be general. It is unclear whether doctors believed the advertising or were induced by the promotions. In either case, prescriptions flowed.

Purdue and other pharmaceutical manufacturers were behaving like stereotypical amoral profit-maximizing companies, but they also met little resistance. National Academies of Sciences, Engineering, and Medicine (2021) noted several system-wide failures in the opioid epidemic. Pharmaceutical distributors and dispensers both have legal obligations to watch for diversion of products, but they are also profit-maximizing entities who benefitted from the sale of opioids.

Physicians are the ultimate gatekeepers for prescription medicine, and many of them do behave far more altruistically than any simple *homo economicus*. Yet not all doctors are saints, and some wrote very large numbers of opioid prescriptions. Prescription Drug Monitoring Program data show that in 2011, prescribers in the top 5 percent of the prescribing distribution wrote 58 percent of total prescriptions in Kentucky (Kentucky Cabinet for Health and Family Services 2005–2018), 36 percent in Massachusetts (Massachusetts Department of Public Health 2000–2018), and 40 percent in California (State of California Department of Justice 2009–2018).

Both the Food and Drug Administration and the Drug Enforcement Administration made decisions that enabled the massive increase in opioid prescriptions (de Shazo et al. 2018; Egilman et al. 2019; Kolodny 2020; Office of the Inspector General 2019). The FDA generally requires at least two long-term studies of safety and efficacy in a particular condition before drug approval, but for OxyContin, the primary trial for approval was a two-week trial in patients with osteoarthritis. Even with this limited evidence, the FDA approved OxyContin "for the management of moderate to severe pain where use of an opioid analgesic is appropriate for more than a few days"—with no reference to any particular condition and no limit to short-term use.

In approving subsequent opioids, the FDA sometimes relied on clinical trials where all patients were initially placed on the active opioid ("open label"). Among those who responded favorably and could tolerate the side effects, some were randomized to continue the therapy and others were switched to a placebo. The idea was that the drug would then be tested in efficacy among those for whom it can be tolerated. However, with this design, withdrawal effects from ending opioid use could be interpreted as efficacy of the therapy.

Among economists, "regulatory capture" is a standard explanation for lax oversight. Two examiners involved in OxyContin's approval by the Food and Drug Administration went on to work for Purdue. When the FDA convened an advisory group in 2002 to examine the harms from OxyContin, eight of the ten experts had ties to pharmaceutical firms.

The Drug Enforcement Agency is in charge of monitoring use of opioids—it maintains the ARCOS data (Automated Reports and Consolidated Ordering System) on prescription drug sales—and of approving increases in production quotas. The DEA approved an increase in production quotas for oxycodone and other opioids numerous times, even as the scope of the opioid epidemic became clear. The DEA has also been blamed for being slow to set up a suspicious order system and to shut down suspected diversion of pills.

A number of states enacted and then expanded their prescription drug monitoring programs (Meara et al. 2016), but these changes came some years later than many would have wished. Earlier, states had moved away from monitoring prescriptions—for example, away from the use of triplicate prescribing forms that have been shown to have slowed the growth of opioid use (Alpert et al. 2019). Private insurers as well were also slow to curb the use of opioid medications, for example through formulary restrictions or prior authorization requirements. In a sense, the case of OxyContin reinforces the point that nimble, well-incentivized, profit-seeking companies can often find their way around a slow-moving regulatory apparatus.

The first hints that OxyContin and its later competitors were no safer than earlier opioids appeared early. By 2001, users had learned that crushing time-release tablets would provide access to the full dose of oxycodone at once (National Drug Intelligence Center 2001). Further, the pain-relieving properties of OxyContin seemed shorter than its promoters promised, and the subsequent pain was intense (Van Zee 2009). Indeed, the flood of opioids nationwide made no difference to the reduction in pain between rounds two and four of the Medical Expenditure Panel Study, regardless of the level of pain in period 2 (as we show in detail in the online Appendix). As the pain continued, there was demand for more frequent and larger doses of medication.

Opioid-related deaths followed the patterns of rising opioid prescriptions. As shown in Figure 1, prior to 1999, only the total rate of opioid deaths was recorded in national data. After 1999, the drug overdose deaths are delineated by type of drug. The overall death rate from drug overdoses in the United States was below five per 100,000 in 1990, despite the notorious crack epidemic and what was known at the time as "Heroin Chic"—when popular musicians, such as Kurt Cobain, were known for their heroin use. The growth rate of opioid deaths shows a trend break almost immediately after Oxycontin was introduced in 1996.

Throughout the 2000s, increasing attention was paid to the abuse of opioids. The National Drug Intelligence Center (2001) report noted with alarm that "the Pike County, Kentucky, Coroner reported 19 Oxycontin-related deaths during the calendar year 2000." In July 2001, the *New York Times* published "The Alchemy of Oxycontin" (Tough 2001), which noted that "the earliest reported cases of OxyContin abuse were in rural Maine, rust-belt counties in western Pennsylvania and eastern Ohio and the Appalachian areas of Virginia, West Virginia and Kentucky." Congressional hearings as early as 2001 described increased deaths and pill diversion from OxyContin (US House 2001).

Policy interventions followed but with a lag. In 2003, the Food and Drug Administration sent a letter warning Purdue that "your advertisements thus grossly overstate the safety profile of Oxycontin," and that "failure to respond to this letter may result in regulatory action, including seizure or injunction, without further notice" (Abrams 2003). That same year, the Drug Enforcement Agency released its OxyContin action plan that called for the "rapid reformulation" of Oxycontin to "reduce the abuse of the project, particularly by injection."

Over time, insurers began to restrict access to opioids through their coverage decisions. When state governments set up prescription drug monitoring programs, physicians could then see how many prescriptions their patients were receiving. Altogether, states added 81 new controlled substance laws between 2006 and 2012 (Meara et al. 2016). State and private lawsuits began to target Purdue Pharmaceuticals and other firms in the opioid business. In August 2010, Purdue finally reformatted OxyContin to make it less vulnerable to abuse. Prescriptions fell, especially for new patients (Zhu et al. 2019), and abuse of OxyContin declined (ICER 2017).

Illegal Innovation and Globalization

Stronger drug monitoring programs and reformulating OxyContin did not end the opioid epidemic; rather, they moved demand into the deadlier illegal market. Cicero and Ellis (2015) found that one-third of opioid users switched to other drugs after the reformulation of OxyContin, and that 70 percent of those who identified an alternative drug specified heroin. The change was greater in those states that had the highest levels of OxyContin misuse prior to 2010 (Alpert, Powell, and Pacula 2018). Figure 1 shows no reduction in overall death rates after 2011; rather, the decline in deaths associated with prescription opioids is fully offset by the increase in deaths from heroin. Indeed, Evans, Lieber, and Power (2019) estimate no impact of the reformulation on overall mortality.

Soon after 2010, imported heroin from Mexico increased in a way that offset the decline in legal opioids. Quinones (2015) describes an almost corporate system, where buyers call dealers, who then deliver heroin on demand. While US demand for heroin generated both competition on quality and violence in Mexico, the US side of the market remained peaceful and efficient, and heroin overdoses rose dramatically until 2016. Better delivery systems that use cell phones also make it easier for consumers to buy heroin without going to potentially dangerous, in-person drug markets.

Over time, heroin was replaced by fentanyl, which is more potent still and is fully synthetic—thus does not need the poppy plant. Fentanyl is far less costly to produce and the drug's concentrated strength means that it is particularly easy to ship in tiny but still powerful quantities. The downside is that it is even more deadly. Even transmission through the skin can kill naïve users, depending on the dose. In an accounting sense, fentanyl-related deaths explain almost all of the increase in drug overdose mortality between 2014 and 2017. Indeed, the ability to access an Asian supplier of fentanyl over the internet may be far more revolutionary than the ability to buy consumer goods on Amazon. We now ask how this is related to the initial addiction to prescription opioids.

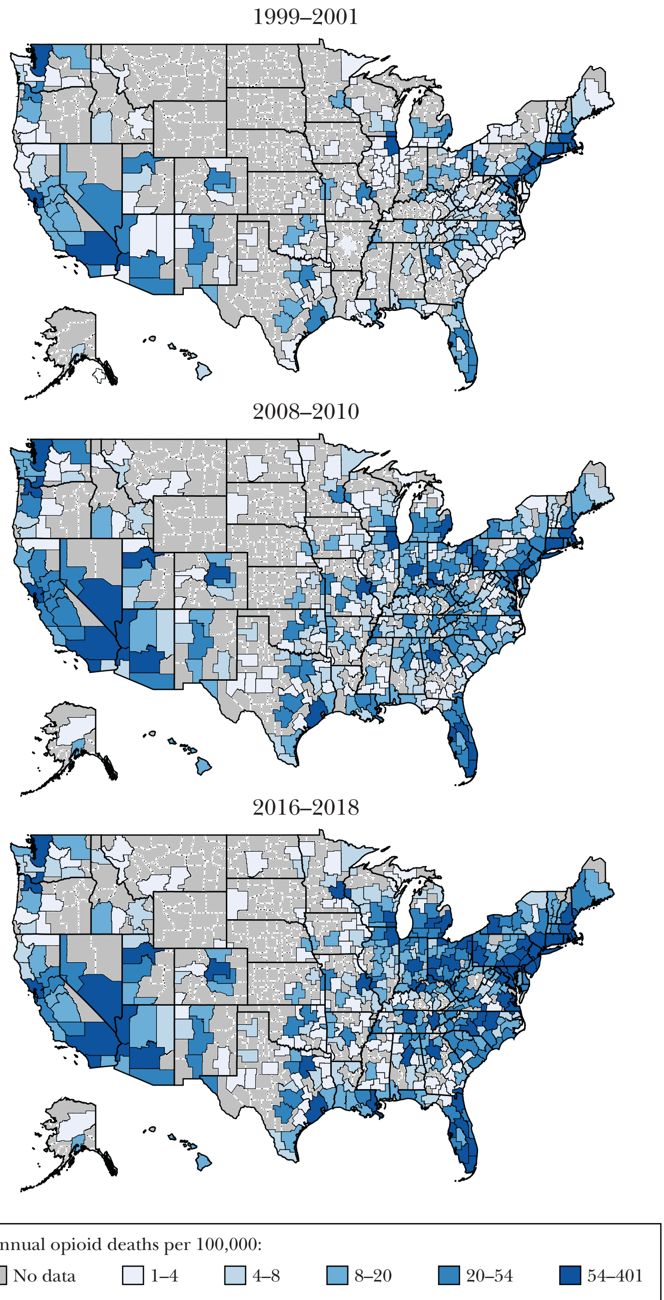
How the Pill Supply Translated into Deaths

The opioid epidemic was not uniform spatially nor was it predominantly an urban phenomenon, unlike crack cocaine in the 1980s (Fryer et al. 2013). Figure 5 shows commuting-zone level maps of opioid deaths in 1999–2001, 2008–2010, and 2016–2018. Up through 2010, death rates rose the most and were highest in Appalachia, the industrial Midwest (Quinones's 2015 book *Dreamland* chronicled southern Ohio), and rural areas of Maine and Nevada. More recently, deaths have increased more in urban areas, as illegal drug markets are more extensive there.

The change in death rates from 2009 to 2017 was more disparate. Generally, areas east of the Mississippi River have higher deaths due to illegal opioids than areas west of the Mississippi River, perhaps related to the type of heroin available prior to the opioid epidemic (Pardo et al. 2019). Fentanyl mixes better with powdered heroin than black tar heroin, and powdered heroin is more common east of the Mississippi. Overall, the cross-area correlation coefficient between opioid-related mortality rates in 2008–2010 and 2016–2018 is 0.82.

To understand the factors explaining these area trends, we use a series of regressions to examine whether the national shift in legal opioid supply had a larger impact on opioid shipments and deaths in communities with more pain. We consider this alongside the alternative hypothesis that opioid use was driven by despair. The data on opioid shipments are from the ARCOS database maintained by the Drug Enforcement Administration. Opioid deaths in each county in each

Figure 5
Age-Adjusted Opioid Deaths (per 100,000) by Commuting Zone



Note: Data are from the Center for Disease Control and Prevention, National Vital Statistics System. Due to CDC restrictions, areas are not shown if the number of deaths < 10 in the three-year total or the commuting zone population is < 100,000 people. These areas are indicated with grey fill and white boundaries. Data are adjusted to the year 2000 population by age and sex, using population data from SEER (NIH 2010) and aggregated from counties to commuting zones using a county to commuting zone crosswalk from the US Department of Agriculture (2019).

year from 1999 to 2018 were obtained through a special request to the Centers for Disease Control and Prevention.

The first column of Table 1 relates opioid shipments in a county to a first measure of pain, the share of people in a county's labor force receiving Social Security Disability Insurance in 1990, before the opioid epidemic, using data from the US Department of Health and Human Services (HHS 1989–1991) and Local Area Unemployment Statistics (LAUS) (Bureau of Labor Statistics 1990). The pain variable is interacted with national opioid shipments; thus, the coefficient indicates how national drug supply translated into pill availability in areas with more and less preexisting pain. The regressions also control for county and year fixed effects. To focus on legal opioids, the regressions are for the period 1997–2010. Pain is a potent predictor of opioid shipments. An area with one standard deviation more people on disability insurance received 23 percent more opioids than the average area. Column 2 considers an alternate measure of pain, the share of the population in the 2002–2005 and 2007 Behavioral Risk Factor Surveillance Survey (BRFSS) that reports joint pain (Centers for Disease Control and Prevention 2021b). This measure is available for 331 large counties. The coefficient is similar to that for receipt of Social Security Disability Insurance: that is, an area with one standard deviation more pain received 18 percent more opioids. The third column interacts national opioid availability with the share of people in the county who were dissatisfied or very dissatisfied with life, also drawn from the 2005–2010 Behavioral Risk Factor Surveillance Survey. Opioid supply did not rise more in areas where more people were more dissatisfied. The fourth column interacts opioid availability with a different metric for despair, the percent of people reporting poor mental health in all 30 days during the past month. This variable is also drawn from BRFSS, and we averaged over 2002–2010 (Centers for Disease Control and Prevention 2021b). Opioid supply increased by 38 percent more in areas with one standard deviation more people reporting this extreme form of mental distress The fifth column includes all four variables together; among them, the share of the population experiencing consistently poor mental health is most related to opioid supply.[5]

The regression in column 6 mirrors that in column 5, with the exception that the dependent variable is the prescription opioid mortality rate. The sample is again 1997–2010. The coefficients are similar to those in column 5. Receipt of disability insurance, joint pain, and extremely poor mental health interact with national opioid supply in leading to prescription opioid deaths; area life dissatisfaction does not interact in this way.

To examine how these factors explain deaths due to illegal opioids, column 7 relates the county's death rate from illegal opioids to the national death rate from illegal opioids interacted with the same county characteristics using data from 2008 to 2017. We also allow for an interaction between deaths from illegal opioids nationally and per capita oxycodone shipments in 2008, to examine whether greater

[5] In the online Appendix, we impute joint pain and life dissatisfaction to all counties, using a LASSO regression in the counties where the data are available. We then relate opioid shipments to predicted pain and despair. The results are similar to what we report in the text.

Table 1
Impact of Drug Shipments Interacted with Pain and Despair on Local Areas

	Interaction with opioid shipments						*Interaction with national illicit death rate*
	Prescription opioid shipments (1997–2010)					*Prescription opioid death rate (1997–2010)*	*Illicit opioid death rate (2008–2017)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction between national opioid shipments/illegal deaths and							
Pain							
Percent of labor force claiming DI (1990)	86.10***				55.90	1.11*	3.02*
	(10.73)				(41.86)	(0.60)	(1.57)
Self-reported joint pain prevalence		65.83**			38.83	0.61**	3.01***
		(31.26)			(32.22)	(0.28)	(1.00)
Despair							
Share dissatisfied/very dissatisfied w/life			33.98		–70.94**	–0.49	1.32
			(28.26)		(29.63)	(0.33)	(1.08)
Extreme mental distress (30 days w/poor mental health)				140.86***	149.03***	1.27***	–0.09
				(28.27)	(36.38)	(0.38)	(1.04)
Opioid shipments							
Oxycodone MME per capita, 1997–2010							4.47***
							(1.11)
Unadjusted mean	368.58	373.68	373.68	373.68	373.75	3.533	8.865
R^2	0.750	0.713	0.707	0.738	0.747	0.448	0.607
Observations	42,966	4,634	4,634	4,634	4.634	3,968	3,641

Note: The dependent variable is age- and sex-adjusted rate in the county and year. Column 1 includes all counties. Columns 2–6 only include counties for which data on joint pain prevalence or share dissatisfied with life were available. National opioid shipments included shipments of oxycodone, hydrocodone, hydromorphone, codeine, morphine, and fentanyl base and were scaled by the change in shipments between 1997 and 2010 in Figure 2. The coefficients represent the impact of one standard deviation higher pain (or despair) times the change in national opioid shipments or national deaths due to illegal opioids which occurred over the period. Counties were weighted by total population in 2005. All regressions control for county and year fixed effects. Standard errors are clustered at the county-level and are reported in parentheses. ****$p < 0.01$ **$p < 0.05$ *$p < 0.10$. For details of the regression data and calculations, see the online Appendix.

prescription opioid use translated into more deaths due to illegal opioids (Evans et al. 2019; Alpert, Powell, and Pacula 2018). Deaths from illegal opioids were greater in areas where pain was greater and where OxyContin shipments were more prevalent. In contrast, despair is not related to deaths from illegal opioids. Overall, therefore, the opioid epidemic was worse in areas where there was more preexisting pain, poor mental health, and where opioid shipments were greater.

Pain is correlated with education at the individual and area level. In the 331 counties with data on pain, the correlation between pain and share with a college degree is –0.11. Pain is also higher in rural areas, areas with more manual labor, and areas with higher rates of obesity. This, in part, explains why opioid death rates are higher in areas with historical manufacturing employment and areas with obesity-related health problems.

Conclusion

America's battle with opioids is not over. A movement to aid a population suffering from chronic pain has become a national crisis, with pain, despair, and entrepreneurship mixed in an unholy brew. The government may assert its authority over legal opioids, but it seems unable to stem most of the current illegal market supply, which increasingly comes in small shipments from Asia or Mexico. Even a pandemic could not slow the deaths from opioid drugs.

In the past, opioid crises ended slowly. Users painfully detoxified when supplies could not be obtained or died after one overdose too many. New users were deterred by fear of addiction and physicians' reluctance to prescribe. The current crisis may follow the same slow and painful path, but that is not entirely clear. New technologies have made it much harder to restrict access to illegal opioids. When the poppy plant had to be grown, supply could be curtailed by eliminating poppy fields. In contrast, fentanyl is much easier to produce, and its sale is much harder to stop. On the other hand, society also has more tools to address addiction—medication-assisted treatment, widespread availability of overdose reversal medication (Narcan), and strong penalties for illegal suppliers—and the use of these treatments is spreading.

Past US public health efforts offer both hope and despair. Nicotine is an extremely addictive substance and yet smoking rates have fallen dramatically over the past five decades, because of both regulation and fear of death. On the other side, the harms of obesity are also well-known and average weights are still increasing. We cannot predict whether opioid addiction will decline like cigarette smoking or persist like obesity.

The medical use of opioids to treat pain will always involve costs and benefits, and the optimal level of opioid prescription is unlikely to be zero. The mistake that doctors and prescribers made in recent decades was to assume overoptimistically that a time release system would render opioids non-addictive. Thousands of years of experience with the fruits of the poppy should have taught that opioids have never been safe and probably never will be.

The larger message of the opioid epidemic is that technological innovation can go badly wrong when consumers, professionals, and regulators underestimate the downsides of new innovations and firms take advantage of this error. Typically, consumers can experiment with a new product and reject the duds, but with addiction, experimentation can have permanent consequences.

■ *We are grateful to Toby Chaiken, Julia Dennett, and Travis Donahoe for expert research assistance. We are also grateful to Leonard Young and Netrali Dalvi of the Massachusetts Prescription Drug Monitoring Program for providing assistance with MA's Prescription Monitoring Program (PMP) data, Adam Berrones and Dave Hopkins for providing assistance with Kentucky's All Schedule Prescription Electronic Reporting (KASPER) data, and Gilbert Samia for providing assistance with California's Controlled Substance Utilization Review and Evaluation System (CURES) data. This research was supported by the National Institute on Aging of the National Institutes of Health under award numbers R37AG047312. Cutler is involved in the multi-district litigation regarding opioids as an expert witness to the plaintiff counties suing opioid manufacturers, distributors, and dispensers. Additional detail on the data and results presented are available in the online Appendix available with this paper at the JEP website.*

References

Abrams, Thomas. 2003. FDA Warning Letter to Purdue http://wayback.archive-it.org/7993/20170112065652/http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/EnforcementActivitiesbyFDA/WarningLettersandNoticeofViolationLetterstoPharmaceuticalCompanies/UCM168946.pdf.

Agency for Healthcare Research and Quality, US Department of Health & Human Services. 2021. "Medical Expenditure Panel Survey (MEPS) Data, 1996-2017." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], 2021-08-13. https://doi.org/10.3886/E146344V1.

Alpert, Abby E., William N. Evans, Ethan M.J. Lieber, and David Powell. 2019. "Origins of the Opioid Crisis and Its Enduring Impacts." NBER Working Paper 26500.

Alpert, Abby, David Powell, and Rosalie Liccardo Pacula. 2018. "Supply-Side Drug Policy in the Presence of Substitutes: Evidence from the Introduction of Abuse-Deterrent Opioids." *American Economic Journal: Economic Policy* 10 (4): 1–35.

Arteaga, Carolina, and Victoria Barone. 2021. "The Opioid Epidemic: Causes and Consequences." Unpublished.

Austin, Benjamin, Edward Glaeser, and Lawrence Summers. 2018. "Jobs for the Heartland: Place-Based Policies in 21st-Century America." *Brookings Papers on Economic Activity* 1970 (1): 151–255.

Avicenna. 1998. *The Canon of Medicine.* Translated by H.A. Hameed. New Delhi: Jamia Hamdard Printing Press. Originally published as *al-Qanun fi al-tibb.*

Becker, Gary, and Kevin Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.

Bentham, Jeremy. 1789. *Introduction to the Principles and Morals of Legislation.* London: Printed for T. Payne and Son, at the Mews Gate.

Blanchflower, David G., and Andrew J. Oswald. 2020. "Trends in Extreme Distress in the United States, 1993–2019." *American Journal of Public Health* 110 (10): 1538–44.

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King, and Kari C.W. Williams. 2019. IPUMS Health Surveys: National Health Interview Survey, Version 6.4 [dataset]. Minneapolis: IPUMS. https://doi.org/10.18128/D070.V6.4.

Booth, Martin. 1996. *Opium: A History.* London: Simon and Schuster.

Brim, Orville Gilbert, Paul B. Baltes, Larry L. Bumpass, Paul D. Cleary, David L. Featherman, William R. Hazzard, Ronald C. Kessler et al. 1995–1996. "Midlife in the United States (MIDUS 1), 1995–1996 (ICPSR 2760)." Inter-university Consortium for Political and Social Research, 2020-09-28. https://doi.org/10.3886/ICPSR02760.v19.

Bureau of Labor Statistics (BLS). 1990. "Local Area Unemployment Statistics: Labor Force Data By County, 1990 Annual Averages." https://www.bls.gov/lau/laucnty90.xlsx.

Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112(49): 15078–83.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity*: 397–476.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism.* Princeton: Princeton University Press.

Case, Anne, Angus Deaton, and Roger Stone. 2020. "Decoding the Mystery of American Pain Reveals a Warning for the Future." *PNAS* 117 (40) 24785–89.

Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Population Health. 2021a. "Behavioral Risk Factor Surveillance System (BRFSS) Annual Survey Data, 1999–2019." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], 2021-08-13. https://doi.org/10.3886/E146342V2.

Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Population Health. 2021b. "SMART: BRFSS City and County Data (2002–2015)." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], 2021-08-13. https://doi.org/10.3886/E146343V1.

Centers for Disease Control and Prevention, National Vital Statistics System. 1990–2018. "Deaths (Mortality)–Multiple Cause of Death, States, and All Counties–Detailed." *National Center for Health Statistics*. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm.

Centers for Disease Control and Prevention, National Center for Health Statistics. 1999–2019. "Multiple Cause of Death." CDC Wonder Online Database, released in 2020. Data are from the Multiple Cause of Death Files. http://wonder.cdc.gov/mcd-icd10.html (accessed August 16, 2021).

Centers for Medicare and Medicaid Services. 2017. "Opioid Oral Morphine Milligram Equivalent (MME) Conversion Factors." https://web.archive.org/web/20200622005723/https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-March-2015.pdf.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2019a. "The Transformation of Manufacturing and the Decline in U.S. Employment." In *NBER Macroeconomics Annual 2018,* Vol. 33, edited by Martin Eichenbaum and Jonathan A. Parker. Chicago: University of Chicago Press.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2019b. "Replication kit for: The Transformation of Manufacturing and the Decline in US Employment." NBER Macroeconomics Annual. https://www.nber.org/books-and-chapters/nber-macroeconomics-annual-2018-volume-33/transformation-manufacturing-and-decline-us-employment.

Cicero Theodore J., and Matthew S. Ellis. 2015. "Abuse-Deterrent Formulations and the Prescription Opioid Abuse Epidemic in the United States: Lessons Learned from OxyContin." *JAMA Psychiatry* 72 (5): 424–30.

Dai, Bingham. 1937. *Opium Addiction in Chicago: A Dissertation.* Shanghai: Commercial Press.

Daly, James R.L. 1900. "A Clinical Study of Heroin." *Boston Medical and Surgical Journal* (142): 190–92.

DeShazo, Richard D., McKenzie Johnson, Ike Eriator, and Kathryn Rodenmeyer. 2018. "Backstories on the US Opioid Epidemic. Good Intentions Gone Bad, an Industry Gone Rogue, and Watch Dogs Gone to Sleep." *American Journal of Medicine* 131(6): 595–601.

Drug Enforcement Administration. 1999–2017. "Automation of Reports and Consolidated Orders System (ARCOS)." https://www.deadiversion.usdoj.gov/arcos/retail_drug_summary/.

Egilman, David S. Gregory B. Collins, Julie Falender, Naomie Shembo, Ciara Keegan, and Sunil Tohan. 2019. "The Marketing of OxyContin: A Cautionary Tale." *Indian Journal of Medical Ethics* 4 (3): 183–93.

Evans, William N., Ethan M.J. Lieber, and Patrick Power. 2019. "How the Reformulation of OxyContin Ignited the Heroin Epidemic." *The Review of Economics and Statistics* 101 (1): 1–15.

Fingerhut, Lois A., and Christine C. Cox. 1998. "Poisoning Mortality, 1985–1995." *Public Health Reports* 113 (3): 218–33.

Flood, S., M. King, R. Rodgers, S. Ruggles, J.R. and Warren. 2020. "Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]." Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D030.V8.0.

Fryer, Roland G., Paul S. Heaton, Steven D. Levitt, and Kevin M. Murphy. 2013. "Measuring Crack Cocaine and Its Impact." *Economic Inquiry* 51 (3): 1651–81.

Gable, Robert S. 2004. "Comparison of Acute Lethal Toxicity of Commonly Abused Psychoactive Substances." *Addiction* 99 (6): 686–96.

Gallup. 1997–2015. "Satisfaction With Personal Life." https://news.gallup.com/poll/1672/satisfaction-personal-life.aspx.

Goodnough, Abby. 2021 "Overdose Deaths Have Surged During the Pandemic, C.D.C. Data Shows." *New York Times,* April 14. https://www.nytimes.com/2021/04/14/health/overdose-deaths-fentanyl-opiods-coronaviurs-pandemic.html.

Grandmaison, Romain LeCour, Nathaniel Morris, and Benjamin T. Smith. 2019. "The US Fentanyl Boom and the Mexican Opium Crisis: Finding Opportunities amidst Violence?" Unpublished.

Hedegaard, Holly, Arialdi M. Miniño, and Margaret Warner. 2018. "Drug Overdose Deaths in the United States, 1999–2017." NCHS Data Brief 329.

Ho, Jessica Y. 2017. "The Contribution of Drug Overdose to Educational Gradients in Life Expectancy in the United States, 1992–2011." *Demography* 54 (3): 1175–1202.

Hoyert, Donna L., Elizabeth Arias, Betty L. Smith, Sherry L. Murphy, and Kenneth D. Kochanek. 2001. "Deaths: Final Data for 1999." *National Vital Statistics Reports.* 49 (8): 1–114.

Institute for Clinical and Economic Review. 2017. *Abuse-Deterrent Formulations of Opioids: Effectiveness and Value.* Boston: Institute for Clinical and Economic Review.

Institute of Medicine (IOM). 2011. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research.* Washington, DC: The National Academies Press.

Jones, John. 1701. *The Mystery of Opium Reveal'd.* London: Printed for Richard Smith at the Angel and Bible without Temple-Bar.

Kentucky Cabinet for Health and Family Services. 2005–2018. "Kentucky All Schedule Prescription Electronic Reporting." (accessed August 18, 2020).

Kochanek, Kenneth D., Elizabeth Arias, and Brigham A. Bastian. 2016. "The Effect of Changes in Selected Age-Specific Causes of Death on Non-Hispanic White Life Expectancy between 2000 and 2014." NCHS Data Brief 250.

Kolodny, Andrew. 2020. "How FDA Failures Contributed to the Opioid Crisis." *AMA Journal of Ethics* 22 (8): 743–50.

Krueger, Alan B. 2017. "Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate." *Brookings Papers on Economic Activity* (2): 1–87.

Maclean, Johanna Catherine, Justine Mallatt, Christopher J. Ruhm, and Kosali Simon. 2020. "Economic Studies on the Opioid Crisis: A Review." NBER Working Paper 28067.

Massachusetts Department of Public Health. 2000–2018. "Prescription Monitoring Program Data." (accessed January 17, 2020).

Massachusetts State Board of Health. 1872. *Third Annual Report of the State Board of Health of Massachusetts (1871).* Massachusetts: State Board of Health.

Meara, Ellen, Jill R. Horwitz, Wilson Powell, Lynn McClelland, Weiping Zhou, A. James O'Malley, and Nancy E. Morden. 2016. "State Legal Restrictions and Prescription-Opioid Use among Disabled Adults." *New England Journal of Medicine* 375: 44–53.

Miniño, Arialdi M., Robert N. Anderson, Lois A. Fingerhut, Manon A. Boudreault, and Margaret Warner. 2006. "Deaths: Injuries, 2002." *National Vital Statistics Reports.* 54 (10): 1–125.

Nahin, Richard L., Bryan Sayer, Barbara J. Stussman, and Termeh M. Feinberg. 2019. "Eighteen-Year Trends in the Prevalence of, and Health Care Use for, Noncancer Pain in the United States: Data from the Medical Expenditure Panel Survey." *Journal of Pain* 20 (7): 796–809.

National Academies of Sciences, Engineering, and Medicine. 2021. *High and Rising Mortality Rates Among Working-Age Adults.* Washington, DC: The National Academies Press.

National Drug Intelligence Center. 2001. Information Bulletin: Oxycontin Diversion and Abuse. https://www.justice.gov/archive/ndic/pubs/651/index.htm.

NIH National Cancer Institute. 2021. "NIH SEER U.S. Population Data - 1969–2019." American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-08-13. https://doi.org/10.3886/E146341V1.

Nutt, David, Leslie King, William Saulsbury, and Colin Blakemore. 2007. "Development of a Rational Scale to Assess the Harm of Drugs of Potential Misuse." *Lancet* 369 (9566): 1047–53.

Office of the Inspector General, Department of Justice. 2019. *Review of the Drug Enforcement Administration's Regulatory and Enforcement Efforts to Control the Diversion of Opioids.* Washington, DC: Department of Justice, Evaluation and Inspection Division 19–05.

Ohler, Norman. 2017. *Blitzed: Drugs in the Third Reich.* Boston: Houghton Mifflin Harcourt.

Pardo, Bryce, Jirka Taylor, Jonathan P. Caulkins, Peter Reuter, Beau Kilmer, and Bradley D. Stein. 2019. *The Future of Fentanyl and Other Synthetic Opioids.* Santa Monica: RAND Corporation.

Pierce, Justin R., and Peter K. Schott. 2020. "Trade Liberalization and Mortality: Evidence from US counties." *American Economic Review: Insights* 2 (1): 47–64.

Podolsky, Scott H., David Herzberg, and Jeremy A. Greene. 2019. "Preying on Prescribers (and Their Patients): Pharmaceutical Marketing, Iatrogenic Epidemics, and the Sackler Legacy." *New England Journal of Medicine* 380: 1785–87.

Porter Jane, and Hershel Jick. 1980. "Addiction Rare in Patients Treated with Narcotics." *New England Journal of Medicine* 302 (2): 123.

Putnam, Robert D. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy* 6 (1): 65–78.

Quinones, Samuel. 2015. *Dreamland*. New York: Bloomsbury Press.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., Sobek, M. 2021. "IPUMS USA: Version 11.0" [dataset]. Minneapolis, MN. https://doi.org/10.18128/D010.V11.0.

Ruhm, Christopher J. 2018. "Corrected US Opioid-Involved Drug Poisoning Deaths and Mortality Rates, 1999–2015." *Addiction* 113 (7): 1339–44.

Ruhm, Christopher J. 2019. "Drivers of the Fatal Drug Epidemic." *Journal of Health Economics* 64: 25–42.

Ryff, Carol, David M. Almeida, John Ayanian, Deborah S. Carr, Paul D. Cleary, Christopher Coe, Richard Davidson et al. 2004–2006. "Midlife in the United States (MIDUS 2), 2004–2006 (ICPSR 4652)." Inter-university Consortium for Political and Social Research, 2017-11-20. https://doi.org/10.3886/ICPSR04652.v7.

Ryff, Carol, David Almeida, John Ayanian, Neil Binkley, Deborah S. Carr, Christopher Coe, Richard Davidson et al. 2013–2014. "Midlife in the United States (MIDUS 3), 2013–2014 (ICPSR 36346)." Inter-university Consortium for Political and Social Research, 2019-04-30. https://doi.org/10.3886/ICPSR36346.v7.

Sarpatwari, Ameet, Michael S. Sinha, and Aaron S. Kesselheim. 2017. "The Opioid Epidemic: Fixing a Broken Pharmaceutical Market." *Harvard Law and Policy Review* 11 (2): 463–84.

Saunders, Nicholas J. 2014. *The Poppy: A History of Conflict, Loss, Remembrance, and Redemption*. London: Oneworld Publications.

State of California Department of Justice. 2009–2018. "Controlled Substance Utilization Review and Evaluation." (January 7, 2021).

Tough, Paul. 2001. "The Alchemy of Oxycontin." *New York Times Magazine*, July 29. https://www.nytimes.com/2001/07/29/magazine/the-alchemy-of-oxycontin.html.

United Nations Office on Drugs and Crime. 2010. *A Century of International Drug Control*. https://www.unodc.org/documents/data-and-analysis/Studies/100_Years_of_Drug_Control.pdf.

United States House of Representatives, Committee on Energy and Commerce, Subcommittee on Oversight and Investigations. 2001. *OxyContin: Its Use and Abuse*. Washington, DC: US Government Printing Office.

Van Zee, Art. 2009. "The Promotion and Marketing of OxyContin: Commercial Triumph, Public Health Tragedy." *American Journal of Public Health* 99: 221–27.

US Census Bureau. 2010a. "2010 ZCTA to County Relationship File." https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2010.html#par_textimage_674173622.

US Department of Agriculture (USDA). 2019. "Commuting Zones and Labor Market Areas." See the data set "2000 commuting zones.xlsx": https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/.

US Department of Health and Human Services (HHS). 1989–1991. "OASDI Beneficiaries by State and County / U.S." Archived at: https://babel.hathitrust.org/cgi/pt?id=uc1.l0078299997&view=1up&seq=5&skin=2021.

Volkow, Nora D. 2014. "America's Addiction to Opioids: Heroin and Prescription Drug Abuse. Senate Caucus on International Narcotics Control, 14, 1–16." See data published in Figure 1: https://archives.drugabuse.gov/testimonies/2014/americas-addiction-to-opioids-heroin-prescription-drug-abuse.

Woolf, Steven H., and Heidi Schoomaker. 2019. "Life Expectancy and Mortality Rates in the United States, 1959–2017." *JAMA* 322 (20): 1996–2016.

Zhu, Wenjia, Michael E. Chernew, Tisamarie B. Sherry, and Nicole Maestas. 2019. "Initial Opioid Prescriptions among U.S. Commercially Insured Patients, 2012–2017." *New England Journal of Medicine* 380: 1043–52.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 197–222

Neighborhoods Matter: Assessing the Evidence for Place Effects

Eric Chyn and Lawrence F. Katz

In 1966, Dorothy Gautreaux and three other Chicago residents sued the Chicago Housing Authority (CHA) in the first major public housing desegregation lawsuit in the United States. Their case highlighted the fact that only 63 of the more than 10,000 public housing units the CHA had built in the previous decade were outside of low-income and racially segregated areas. The resulting settlement reached in 1976 between the CHA and the US Department of Housing and Urban Development created the Gautreaux Assisted Housing Program, which provided housing vouchers and mobility assistance to a limited number of African-American residents from public housing projects in Chicago. From 1976 to 1998, the Gautreaux program helped around 7,100 families move to private rental housing in areas ranging from inner-city neighborhoods to upper-middle-class suburbs.

The Gautreaux program provided a promising opportunity for researchers to study the importance of neighborhoods. Social scientists have long hypothesized that living in a disadvantaged area directly affects the outcomes of adults and life courses of children. Descriptive research has supported this idea by showing that individuals living in high-poverty areas fare worse both contemporaneously and over the long-run in terms of important outcomes such as education, criminal involvement, health, and earnings (Wilson 1987; Jencks and Mayer 1990; Brooks-Gunn

■ *Eric Chyn is an Assistant Professor of Economics, Dartmouth College, Hanover, New Hampshire. Lawrence F. Katz is the Elisabeth Allison Professor of Economics, Harvard University, Cambridge, Massachusetts. Chyn is a Faculty Research Associate and Katz is a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are Eric.T.Chyn@Dartmouth.edu and lkatz@harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.197.

et al. 1993; Sampson, Morenoff, and Gannon-Rowley 2002; Sharkey and Faber 2014). Yet a lingering concern is that such correlations between individual outcomes and neighborhood characteristics may reflect unmeasured differences in individual characteristics that affect both outcomes and selection into living in a disadvantaged area. After all, under everyday circumstances, a household's decision concerning where to live is not random. Prior observational studies suggest that the magnitude of estimated neighborhood effects is often reduced when researchers control for detailed family background measures (Solon 1999).

A crucial feature in the Gautreaux setting was what appeared to be a substantial degree of randomness in the process by which households were matched to available rental units in different neighborhoods. Housing counselors identified rental units—regardless of whether the units were in the city or a suburb of Chicago—and offered them to families based on their position on a waiting list. This process potentially limited the extent to which difficult-to-measure household advantages and disadvantages drove the neighborhood choices of Gautreaux families. Indeed, influential early work on Gautreaux strongly suggested that place of residence mattered: families who moved to the more advantaged suburbs had better outcomes. Popkin, Rosenbaum, and Meaden (1993) studied a survey of female household heads from Gautreaux, finding substantially higher employment rates for the suburban movers compared with their counterparts who stayed in the city. Children from Gautreaux households that moved to the suburbs were less likely to drop out from high school, were more likely to attend a four-year college, and had higher rates of employment relative to those whose families moved within the city of Chicago (Kaufman and Rosenbaum 1992; Rosenbaum 1995).

But the Gautreaux results became less clear as further evidence accumulated. When Mendenhall, DeLuca, and Duncan (2006) conducted a longer-term and more comprehensive follow-up analysis, they found little systematic effects of suburban (versus city) relocation on employment or welfare receipt for the female household heads. Their work and subsequent studies provide evidence that the placement type (specifically, a suburban or city rental unit) was systematically related to pre-move household characteristics, suggesting that the Gautreaux setting may not approximate a randomized experiment (Votruba and Kling 2009; Deluca et al. 2010).

Inspired by the early promising findings from Gautreaux as well as by its methodological limitations, the subsequent generation of neighborhood-effects studies have addressed selection bias by relying on randomized field experiments and on quasi-experimental sources of variation. For example, several studies use data from the US Department of Housing and Urban Development's Moving to Opportunity randomized housing mobility demonstration, which helped a treatment group of public housing families move to lower-poverty areas by providing them with housing vouchers and mobility counseling. Other studies have relied on quasi-experimental research designs including comparisons of children who moved to new areas at different ages and examinations of individuals forced to relocate due to plausibly exogenous events such as natural disasters or public housing demolitions.

In this essay, we summarize what has been learned about the causal impact of neighborhoods in the two decades since the early research on Gautreaux. Our discussion is motivated by new findings that have reshaped how scholars understand the importance of neighborhood environments for adults and children. We concentrate on empirical studies and do not attempt a comprehensive review of the methodological and econometric issues covered in prior reviews such as Durlauf (2004) and Graham (2018). Our work complements and extends other recent reviews of neighborhood-effects research such as Sharkey and Faber (2014) and Galster and Sharkey (2017). Although our focus is on evidence from high-income countries, we believe that lessons regarding neighborhood effects in developed countries may have relevance for understanding neighborhood influences in developing countries as well.

We begin with descriptive evidence indicating that key outcomes of adults and children are strongly correlated with neighborhood poverty rates. Such patterns have motivated the search for compelling approaches to estimate the causal effects of neighborhoods on a range of outcomes. We then sketch a conceptual model that highlights the potential influences of *current* neighborhoods through contemporaneous (or situational) effects and of *past* neighborhoods through exposure (or developmental) effects that accumulate during childhood. The hypothesis that neighborhood effects on socioeconomic and health outcomes operate through the length of exposure to different neighborhood environments has been emphasized by Wilson (1987), Jencks and Mayer (1990), and Sampson (2012). A key prediction of the exposure hypothesis is that the gains to moving to neighborhoods with beneficial attributes will be larger for children who are younger at the time of a move and thus exposed for a longer period.

Next, we discuss the findings from recent experimental and quasi-experimental studies for adults and children separately. Our review of the evidence can be summarized in two main points. First, the findings for adults require a nuanced interpretation. Recent work provides little evidence that changing neighborhoods *within* a city (or commuting zone) has impacts on contemporaneous economic outcomes (at least for adult heads of low-income households). The within-city pattern contrasts with several studies that show longer-distance relocations by adults to higher-wage labor markets (cities or regions) can improve their economic outcomes. The evidence for health outcomes and health-related behaviors (like smoking) is more consistent and suggests that adults benefit from both local and longer-distance moves to higher-opportunity areas. Second, studies of children strongly support the existence of effects in which longer exposure to "better" neighborhood environments during childhood leads to improved long-run outcomes.

We also assess the evidence for different mechanisms that could drive the observed neighborhood impacts. For adults, we discuss how the evidence on the effects of within-city relocation is at least superficially inconsistent with the influential spatial mismatch hypothesis of Kain (1968). The analysis of adult health outcomes suggests that neighborhood stressors and health-related behaviors (like smoking) are key channels. For children, at least five factors appear to be mediators of place effects: school quality, peer influences, pollution, exposure to violence, and criminal justice policies.

Finally, we discuss the implications of recent studies of neighborhood effects for the design of housing policies and conclude by raising outstanding research questions. Open issues include understanding the relative importance of different mechanisms behind neighborhood effects, assessing the general equilibrium impacts of housing mobility policies and other low-income housing assistance programs, and examining the impacts of place-based neighborhood revitalization policies on the preexisting residents of targeted areas.

Background on Neighborhoods and Outcomes

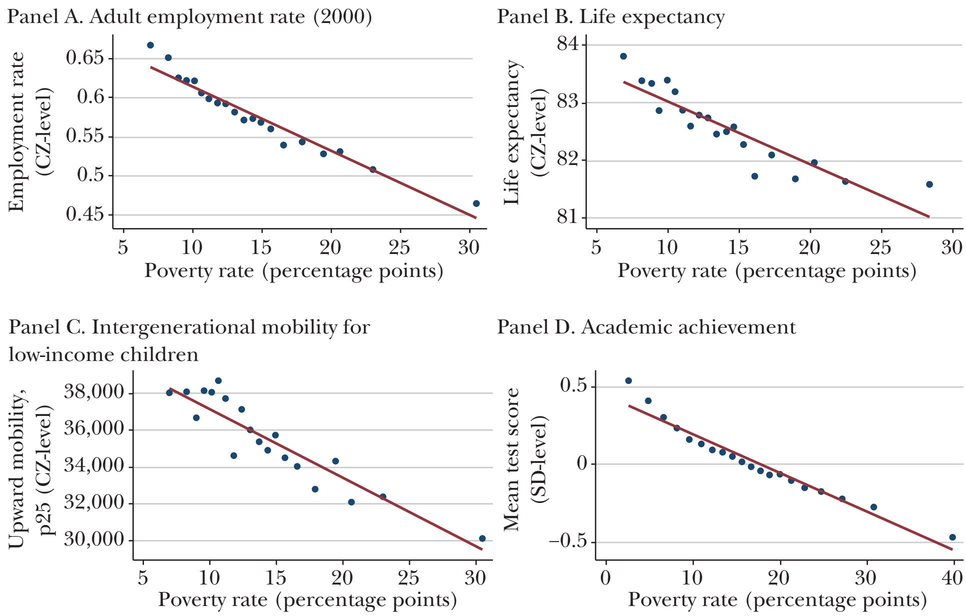
In this section, we use publicly available US data to conduct descriptive analyses that motivate the idea that place of residence matters. For the unit of analysis, we focus on several geographies analyzed in prior studies of neighborhood effects. The largest units of geography that we consider are commuting zones, which are aggregations of counties based on commuting patterns in the 1990 Census and can be viewed as approximating a local labor market. There are 741 US commuting zones. We also examine relationships at more granular levels such as school districts and census tracts. There are over 12,000 school districts and about 72,000 census tracts in the United States. Census tracts typically have a few thousand residents and come closer to what most people commonly refer to as a "neighborhood." To classify these geographic areas by economic opportunity, we use the poverty rate from the 2000 Decennial Census.[1] Poverty rates are a widely used measure in the neighborhood-effects literature (Sampson and Sharkey 2008) and can be broadly interpreted as a summary index of the bundle of characteristics associated with a neighborhood (Kling, Liebman, and Katz 2007).

Figure 1 illustrates a strong association between an area's poverty rate and various outcomes for adults and children. Each panel plots averages based on grouping commuting zones (in panels A, B, and C) or school districts (in panel D) into one of 20 "bins" by poverty rate. The results in panel A show that a one percentage point increase in the poverty rate in a commuting zone is associated with a 0.8 percentage point decline in the adult employment rate using data from the 2000 US Census. Panel B shows that adult health (as measured by life expectancy at age 40) also decreases with the poverty rate. Life expectancy is measured using data from Chetty et al. (2016a) and is based on mortality records from the Social Security Administration.

The results in Figure 1 also show that upward mobility and academic achievement of children both decrease with the poverty rate. The measure of upward mobility is the mean household income (measured at ages 31–37) for children who

[1] For our analysis of school district test scores, we rely on poverty rates using measures from the American Community Survey for the years 2007–2011 and 2012–2016.

Figure 1
Associations between Adult and Child Outcomes and Neighborhood Poverty



Source: Details on data sources are provided in the notes to Table 1.
Note: This figure provides binned scatter plots of the relationship between the poverty rate and the following measures of average resident outcomes: employment rates, life expectancy, upward mobility, and test scores. The unit of analysis in panels A, B, and C is a commuting zone. In panel D, the unit of analysis is a school district.

grew up in each commuting zone and were born to low-income parents (those at the 25th percentile of the income distribution) from the Opportunity Atlas (Chetty et al. 2020a).[2] The measure of achievement is based on the mean of standardized test scores for school districts from the Stanford Education Data Archive. Panels C and D of Figure 1 show that a one percentage point increase in the poverty rate is associated with declines of $371 in a child's expected adult income and 0.025 standard deviations in academic achievement, respectively. All the relationships in Figure 1 are statistically significant at the 1 percent level, as indicated in the regression results reported in columns 1–4 of Table 1.

[2] Chetty et al. (2020a) created the Opportunity Atlas using data from the US Census and federal income tax returns. They study a sample of 20.5 million children born between 1978 and 1983 who are in their mid-30s in 2014–2015. Children are mapped back to Census tracts where they lived until age 23. They construct a measure of average outcomes by measuring parent and child percentile ranks in the national distribution. The Opportunity Atlas contains measures of the average percentile for children in each area. To aid interpretation of this outcome, we converted their rank outcomes into real 2015 US dollars using the national income distribution.

Table 1
Associations between Adult and Child Outcomes and Neighborhood Poverty

Dependent variables:	Adult employment rate (2000) (1)	Life expectancy (2)	Upward mobility, p25 parents (3)	Test-based achievement (4)	Adult employment rate (2000) (5)	Upward mobility, p25 parents (6)
Poverty rate	–0.00821	–0.109	–371.5	–0.0248	–0.00524	–384.3
(percentage points)	(0.000)	(0.011)	(39.907)	(0.0003)	(0.0001)	(9.875)
Observations	741	595	741	12,601	72,416	71,923
Level of analysis	CZ	CZ	CZ	School district	Tract	Tract
Controls	None	None	None	None	County FE	County FE
Mean	0.578	82.58	35,469	0.0245	0.594	34,455
R^2	0.533	0.241	0.096	0.456	0.537	0.539

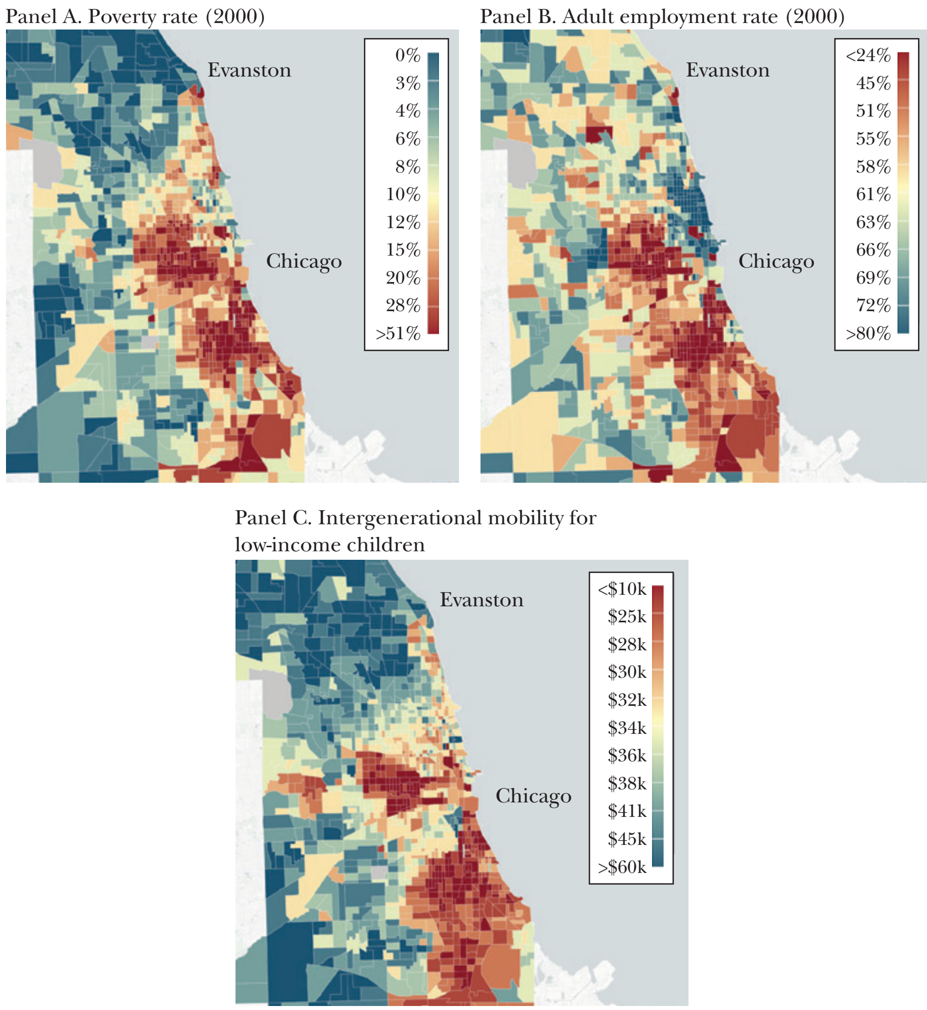
Note: This table reports estimates from a regression model where the dependent variable is a measure of adult or child outcomes (specified in each column header) in a geographic area. Geographic areas are commuting zones (CZ), school districts, or Census tracts. The independent variable of interest is a location specific measure of the poverty rate (the fraction of residents living below the poverty line). Columns 1, 2, 3, 5, and 6 use poverty rates from the 2000 Decennial Census. Column 4 uses poverty rates averaged over 2007–2016 from the American Community Survey (the combined files for the years 2007–2011 and 2012–2016). The dependent variables in columns 1 and 5 are measures from the 2000 Decennial Census. Column 2 uses the life expectancy measure from Chetty et al. (2016a,b) based on mortality data from Social Security Administration death records. Columns 3 and 6 use the "Upward Mobility" measure from the Opportunity Atlas (Chetty et al. 2020a,b) which is the mean later-life household income rank (measured at ages 31–37) for children whose parents were at the twenty-fifth percentile of the national income distribution. Column 4 uses the test-based achievement measure from the Stanford Education Data Archive (SEDA) which is an estimate of mean test scores on a cohort standardized scale. The test score means are constructed using data from the National Assessment of Educational Progress (NAEP) as detailed in Fahle et al. (2019). Standard errors are clustered at the county level in columns 5 and 6.

The correlations between the poverty rate and outcomes are not simply due to broad differences across metropolitan areas. Columns 5 and 6 of Table 1 present correlations between poverty rates and resident outcomes at the census-tract level using data on all US tracts and specifications that control for county fixed effects. These within-county results generate estimates similar to what we observe in the commuting-zone level analysis.

Figure 2 provides another illustration of within-city patterns using Chicago as an example and provides maps of tract-level poverty rates, adult employment, and child outcomes. Dark red indicates areas with relatively worse outcomes (for example, higher poverty) while dark blue areas have better outcomes (like lower poverty). Panel A shows substantial variation in poverty rates within the city. The high poverty tracts are predominately located in the western and southern areas. In line with the results from Table 1, panels B and C show that these high-poverty areas are also those where adults have the lowest employment rates and low-income children have the least upward mobility.

Associations between neighborhood poverty and individual contemporaneous and longer-term outcomes could be driven by two very different sources. One possibility is that neighborhood environments have causal impacts on adults and

Figure 2
Within-City Associations between Adult and Child Outcomes and Neighborhood Poverty, by Census Tract in Chicago



Source: Details on data sources are provided in the notes to Table 1.
Note: This figure provides maps of tract-level poverty rates in 2000 (panel A), mean adult employment rates in 2000 (panel B), and upward mobility for low-income children (panel C) in Chicago.

children. Another possibility is that the observed patterns reflect the non-random sorting of the types of people who end up living in disadvantaged areas. In the next section, we sketch a model of neighborhood effects and use it to illuminate the need for experimental or quasi-experimental approaches to estimate causal neighborhood effects.

Conceptual Framework for Neighborhood Effects

Models in economics and sociology postulate a "production function" relationship in which the outcomes for an individual (such as earnings, health status, or academic achievement) are influenced by both contemporaneous and developmental neighborhood effects in addition to family inputs.[3] Developmental neighborhood effects (also called "exposure effects") are typically hypothesized to depend on the length of past exposure to neighborhoods of different quality, especially during childhood. In addition, neighborhood environments might also have larger and long-lasting impacts at certain ages, like early childhood or the start of adolescence, as embodied in the "critical age effects" hypothesis (Almond and Currie 2011; Heckman and Mosso 2014). Note that neighborhood effects can operate through multiple channels including peer influences, neighborhood safety and exposure to violence, school quality, the physical environment, and access to employment and criminal opportunities (Kain 1968; Wilson 1987; Jencks and Mayer 1990; Glaeser, Sacerdote, and Scheinkman 1996).

A prominent approach to analyzing neighborhood effects is the canonical linear-in-means model of social interactions that features only contemporaneous effects (Manski 1993; Brock and Durlauf 2001). In this model, the impact of neighborhoods stems from three sources. First, endogenous (peer) effects arise due to the propensity for individual behavior to depend on the expected (mean) behavior of their neighborhood peers. Second, exogenous effects represent the possibility that individual behavior is shaped by the average characteristics of neighbors (such as their income and education). Third, correlated effects refer to the fact that individuals within a neighborhood face the same institutional and physical environments including schools, law enforcement policies, and levels of pollution.

In this canonical model, it is typically not possible to identify endogenous effects separately either from exogenous effects or from unmeasured correlated effects (Manski 1993). However, a reduced form can be estimated to provide evidence of contemporaneous neighborhood effects by regressing the outcome of interest on an individual's own characteristics, the mean characteristics of current residential neighbors (like their socioeconomic background), and other current neighborhood characteristics (like school resources). The reduced form estimates can provide suggestive evidence for the presence of peer effects (from the impacts of mean neighbor background characteristics) and effects of neighborhood attributes (from the impacts of specific neighborhood characteristics). Experimental or quasi-experimental variation in peer behavior is needed to estimate causal endogenous peer effects as in the randomized field experiments analyzed by Duflo and Saez (2003) and Bursztyn et al. (2014).

[3]For a formal exposition of the model behind our thinking in this paper and how it leads to some of the prominent equations that are estimated in empirical work, see online Appendix A available with this paper at the *JEP* website. Topa and Zenou (2015) provide a more detailed overview of theoretical models of neighborhood effects.

More recently, much attention has focused on models that center solely on developmental neighborhood effects for children. For example, Chetty and Hendren (2018) study the impacts of moving a child to a new neighborhood where other children typically do well. They characterize neighborhoods by measuring the mean adult outcomes of children who spend their entire childhood in an area ("permanent residents"). Their approach studies children who moved at different ages and examines how their later-life adult outcomes vary with the duration of childhood exposure to more advantaged neighborhoods. The idea that neighborhoods have exposure effects also has been examined in observational studies in sociology (Wodtke, Harding, and Elwert 2011; Sharkey and Faber 2014) and is closely related to models of social capital in economics (Loury 1977; Glaeser, Laibson, and Sacerdote 2002).

Persuasive statistical identification of contemporaneous or developmental neighborhood effects can be challenging due to non-random selection of people into neighborhoods. The neighborhood-effects literature has followed broader trends in economics to address self-selection bias by using experimental and quasi-experimental approaches. The key feature is that the analysts study settings in which there is substantial random (exogenous) variation in exposure to different neighborhood environments. Such an approach is clearest in research using the experimental data from the Moving to Opportunity demonstration. Other studies that rely on quasi-experimental methods address concerns over selection bias by comparing groups where the variation in neighborhood exposure approximates random assignment.

Evidence on Neighborhood Effects for Adults

Beginning in 1994, the Moving to Opportunity housing mobility demonstration randomized access to housing vouchers and assistance in moving to less-distressed communities to about 4,600 families living in public housing projects located in deeply impoverished neighborhoods in five cities: Baltimore, Boston, Chicago, Los Angeles, and New York. The program randomized families into three groups: 1) a low-poverty voucher group (also called the "experimental group") that was offered housing-mobility counseling and restricted housing vouchers that could only be used to move to low-poverty areas (Census tracts with 1990 poverty rates below 10 percent); 2) a traditional voucher group that was offered regular Section 8 housing vouchers that had no additional locational constraints (also called the Section 8 group); and 3) a control group that received no assistance through the program.

The Moving to Opportunity experiment generated large changes in neighborhood environments. One year after baseline, the average adult in the control group was living in a neighborhood with a tract poverty rate of 50 percent. Moving with a low poverty or traditional voucher reduced average tract poverty rates by 35 and 21 percentage points, respectively (Ludwig et al. 2013). At the time of the final follow-up survey conducted 10 to 15 years after random assignment, the Moving to Opportunity groups showed large differences in duration-weighted average neighborhood poverty rates since program entry. Also, families in the experimental voucher group reported feeling safer in their neighborhoods and less likely to have observed conditions of local disorder such as drug activity.

Policymakers had hoped that moves through Moving to Opportunity would generate gains in work and reductions in welfare participation for the adult household heads, but there is little evidence of improved economic self-sufficiency from moves to lower-poverty neighborhoods for adults. For example, Kling, Liebman, and Katz (2007) find no detectable impacts on economic self-sufficiency four to seven years after random assignment, and Ludwig et al. (2013) find a similar pattern 10 to 15 years after random assignment. Chetty, Hendren, and Katz (2016) use administrative Internal Revenue Service records to conduct a longer-term analysis and similarly find no statistically significant effects of Moving to Opportunity vouchers on earnings or employment of adults.

At the same time, the Moving to Opportunity program did generate statistically significant improvements in physical and mental health as well as subjective well-being for adults. Specifically, the low-poverty voucher group experienced a decreased incidence of extreme obesity and diabetes (Ludwig et al. 2011), a reduction in psychological stress and increase in calmness and tranquility (Katz, Kling, and Liebman 2001; Kling, Liebman, and Katz 2007), and an increase in subjective well-being (Ludwig et al. 2012) in the short-run (one to three years after random assignment), interim (four to seven years after random assignment), and final (10 to 15 years after random assignment) follow-up surveys.

Since the launch of the Moving to Opportunity demonstration experiment, studies of natural experiments generated by housing assistance policies have provided further evidence on neighborhood effects for adults.[4] Chyn (2018) examined neighborhood effects for adults by studying public housing demolitions in Chicago during the 1990s. To estimate causal impacts, Chyn compared displaced and non-displaced public housing residents who appear similar before some were displaced by the demolition. This approach revealed that demolition notably improved residential outcomes, as displaced households typically relocated to areas with less poverty and lower crime rates. Despite improvements in neighborhood quality, there were no statistically significant impacts on employment, earnings, or participation in social assistance programs for the displaced adults who had children. In line with these results from Chicago, Haltiwanger et al. (2020) study a broader national set of public housing demolitions and find that displaced adult household heads experience no employment or earnings gains over the next ten years even when moving to neighborhoods with greater nearby job accessibility.

[4] In the context of a low-income country, Barnhardt, Field, and Pande (2017) studied an urban housing lottery in India that provided winning residents of slums with the opportunity to move to improved housing in a new neighborhood. They found that 14 years after the housing allocation, lottery winners lived in safer and cleaner locations, but there was no evidence of improvements in other socioeconomic outcomes such as income or labor force participation.

Recent studies of within-country regional migration provide additional evidence on the importance of place effects for adults.[5] Deryugina, Kawano, and Levitt (2018) and Deryugina and Molitor (2020) use a quasi-experimental approach to estimate impacts of relocation due to Hurricane Katrina on income and health, respectively: specifically, they compare outcomes for displaced New Orleans residents with a comparison group living in similar US cities using a difference-in-difference framework. They find long-run improvements in labor market outcomes and reductions in mortality for the elderly, which were likely driven by the fact that Hurricane Katrina victims typically moved to areas with stronger labor markets and better health outcomes. Collins and Wanamaker (2014) and Boustan (2016) study the Great Migration—the massive movement of African-Americans from the rural South to urban areas in the North—and find large increases in earnings based on a sibling fixed effects approach. Finally, Black et al. (2015) and Johnson and Taylor (2019) study historical US rural–urban migration using instrumental variable strategies. Their results show that moving to urban areas was damaging for health and that this impact may have been mediated by changes in migrant health-related behaviors.

Overall, the experimental and quasi-experimental evidence on adults suggests two main findings. First, relocation *within* a city or commuting zone does not seem to improve earnings or other economic outcomes for adults, but long-distance migration to higher-wage areas or stronger labor markets generates notable economic gains. The significant negative cross-sectional relationship for adult employment and neighborhood poverty within a commuting zone (column 5 of Table 1) appears to largely reflect selection and residential sorting. Place effects on contemporaneous adult labor market outcomes appear to operate at a broader geographic level (the local labor market, commuting zone, region, or country) than one's residential neighborhood within a commuting zone. Second, place of residence has large impacts on physical and mental health outcomes for adults in studies of both within-city and cross-city moves.

Mechanisms for Adults

What do the empirical findings for adults imply for theories of neighborhood effects? The conclusion that within-commuting-zone moves into areas with higher employment have little impact on economic outcomes for low-income minority household heads superficially appears to be inconsistent with the "spatial mismatch

[5] Studies of international immigration also provide evidence on the impacts of place of residence. Abramitzky, Boustan, and Erikkson (2012) estimate large economic returns to immigration from Norway to the United States during the late nineteenth century. Clemens (2013) shows that Indian software workers who win a US visa lottery to immigrate earn substantially more than their coworkers who perform similar duties but lost the lottery. McKenzie, Stillman, and Gibson (2010) find large earnings gains for Tongans who randomly win the opportunity to migrate to New Zealand.

hypothesis" (Kain 1968; Wilson 1987; Holzer 1991).[6] This hypothesis posits that racial economic disparities have resulted in part from unequal access to suburban job opportunities: that is, as a large number of jobs and white residents shifted from urban to suburban areas in the post-1950s period, a combination of housing market discrimination and poor public transportation options limited the access of racial and ethnic minorities to those jobs. A contributing factor could be greater discrimination against minority job applicants by employers in more affluent and mostly white neighborhoods, as found by Agan and Starr (2020) in a recent randomized audit study. An augmented spatial mismatch model potentially consistent with the findings in mobility studies is one in which housing market discrimination coupled with high commuting costs could reduce the labor market options for minority workers, effectively increasing employer monopsony power in the low-wage labor market, and thereby serving to reduce employment rates and wages for less advantaged minorities throughout a metropolitan area.

For adult health, recent studies are consistent with two broad conclusions regarding theory and mechanisms. First, exposure to community disorder and violence has adverse impacts on mental health. Participants in the Moving to Opportunity demonstration stated that concerns about neighborhood violence and crime were the primary motivations for their desire to move out of public housing, and the moves to lower-poverty areas were associated with reductions in neighborhood crime rates and increases in neighborhood collective efficacy. These moves to higher-opportunity neighborhoods also led to lower self-reports of criminal victimization and improved perceptions of neighborhood safety (Katz, Kling, and Liebman 2001; Kling, Liebman, and Katz 2007; Ludwig et al. 2013). Second, place of residence may help to shape health behaviors. For example, Black et al. (2015) and Johnson and Taylor (2019) provide suggestive evidence that historical US rural–urban migration increased mortality due to changes in smoking behavior and alcohol consumption. Moreover, Kling, Liebman, and Katz (2007) find that the reductions in obesity and the mental health gains for the low-poverty voucher treatment group in Moving to Opportunity were modestly associated in the medium-term with increases in physical exercise and dietary shifts toward fruits and vegetables.

Evidence on Neighborhood Effects for Children

In the initial Moving to Opportunity studies, researchers focused on older children and found mixed evidence on the effects of moving to low-poverty neighborhoods. For example, Kling, Liebman, and Katz (2007) found beneficial effects on education, risky behaviors, and physical health for female youth in the five-year post-enrollment survey. However, the beneficial impacts for teenage girls stood in

[6] In contrast, the finding of Haltiwanger et al. (2020) that labor market outcomes at age 26 for children displaced in public housing demolitions are positively related to neighborhood job accessibility is consistent with the traditional spatial mismatch hypothesis playing a role for young minority adults.

contrast to adverse effects of moving for teenage boys. To study these gender differences in effects, Clampet-Lundquist et al. (2011) collected in-depth interviews for a subsample of children in the experiment. They found that gender differences in daily routines, the ability to fit in with neighborhood norms, and neighborhood navigation strategies may have contributed to how girls appeared to benefit more than boys from moves to lower-poverty neighborhoods.

Only recently has enough time passed to study long-run outcomes for the younger children in the Moving to Opportunity demonstration. Chetty, Hendren, and Katz (2016) linked the Moving to Opportunity sample to administrative tax records to study impacts for children of all ages. A major goal of their analysis was to study whether the duration of childhood exposure to new neighborhood environments matters. They do this by comparing program impacts on younger children (those younger than age 13 at random assignment) to older children (those who were 13 to 18 years old at random assignment). They found substantial positive effects on adult earnings and the likelihood of attending college for younger children in the experimental treatment group.[7] However, Chetty, Hendren, and Katz found no detectable effects (or even negative impacts) on longer-run measures of adult earnings and college attendance for older children in the experimental treatment group. The long-run economic gains from moves to lower-poverty areas for the younger children and, if anything, the adverse effects for the older children in Moving to Opportunity are apparent and similar in magnitude for both male and female children, despite the observed gender differences in short-run adjustments to new neighborhood environments.

The differential pattern of treatment effects on economic outcomes for younger and older children in the Moving to Opportunity experiment are consistent with a model that contains a disruption cost for moving to a different type of neighborhood and allows the benefits from relocating to lower poverty areas to be proportional to the duration of exposure during childhood. Exposure effects can be sufficiently large to outweigh a disruption cost when moves occur at early ages. Disruption effects may occur because moving to a different environment in childhood, especially during adolescence, could have negative impacts on social ties or other adverse effects on development (Wood et al. 1993).

Chyn (2018) provides further evidence on the importance of childhood neighborhoods by exploiting the vagaries of timing and choice of units impacted by public housing demolitions. He finds that public housing demolitions in Chicago led to relocations to lower-poverty neighborhoods and significantly improved the later-life labor market and criminal justice outcomes for children.[8] Notably, he found

[7] Other studies have considered long-run impacts of Moving to Opportunity on additional outcomes. Pollack et al. (2019) find that moving led to reductions in annual hospital spending for younger children. Miller and Soo (2021) detect increases in credit scores and credit use for those making Moving to Opportunity moves as young children.

[8] Additional work by Chyn and Haggag (2020) shows that children displaced by public housing demolitions were much more likely to be politically active relative to their non-displaced counterparts.

the long-run positive impacts are larger for children displaced before age 13.[9] In a similar vein, Nakamura, Sigurdsson, and Steinsson (2021) study long-term effects of forced relocation due to a 1973 volcanic eruption in Iceland. They find increased earnings and educational outcomes for the displaced, but only for those who were younger than age 25 at the time of the eruption. Overall, the results from these two quasi-experimental settings are in line with the childhood exposure evidence in Chetty, Hendren, and Katz (2016).

Some of the most compelling evidence of neighborhood exposure effects for children can be found in Chetty and Hendren (2018). They use quasi-experimental methods and tax records to study US children (born from 1980 to 1988) from seven million families that moved across commuting zones. They compared outcomes for children whose families moved when they were different ages to understand how effects vary with the duration of time spent living in more-advantaged areas (those where the children of permanent residents have higher average outcomes). Their approach identifies childhood neighborhood exposure effects under the assumption that selection effects do not vary with the child's age at move among families moving across the same types of locations. Chetty and Hendren used three primary approaches to support the reasonableness of this assumption: 1) their results are robust to controlling for family fixed effects (thereby relying only on the variation in neighborhood exposure among siblings); 2) their findings are robust to controlling for time-varying observable family characteristics reflecting changes in parental economic circumstances; and 3) they obtain similar results when focusing on a subset of moves that are likely to be driven by plausibly exogenous regional shocks such as natural disasters.

The findings of Chetty and Hendren (2018) reveal significant neighborhood exposure effects on intergenerational mobility: the adult incomes of children who moved converge to the adult incomes of children of permanent residents in the destination at a rate of 4 percent per year of childhood exposure. In other words, this estimate suggests that a young child who moves at birth to a better area and stays there for 15 years would pick up 60 percent of the difference in permanent resident adult economic outcomes between their origin and destinations. Additional work has shown that there are similarly large exposure effects for other long-term outcomes such as college attendance, marriage, teenage birth rates, and incarceration (Chetty and Hendren 2018; Chetty et al. 2020a).

Several recent papers have used the same empirical framework from Chetty and Hendren (2018) and replicated their findings using data from other countries. Deutscher (2020) finds evidence of exposure effects on labor market outcomes using tax records from Australia. Notably, his analysis estimated exposure effects

[9] Jacob (2004) similarly provides evidence of the short-run effects on children of public housing demolitions in Chicago. He used data from Chicago Public Schools and found no beneficial impacts of relocation due to demolition on high school graduation or academic achievement. Haltiwanger et al. (2020) study a national sample of public housing demolitions and find large long-run labor market gains from public housing demolition for children at ages 10 to 18 years at the time of the demolition.

from infancy onward and showed that exposure effects in his setting are largest in the teenage years. Laliberté (2021) detect exposure effects on educational attainment in Canada using academic records from Montreal. His estimates are of similar magnitude in Chetty and Hendren (2018): educational attainment of movers converges linearly at a rate of 4.5 percent per year.

Broadly speaking, these recent experimental and quasi-experimental studies provide robust evidence that childhood neighborhood of residence matters for long-run outcomes.[10] In particular, the effects of moving at early ages to more advantaged neighborhoods seem positive for labor market outcomes (Chetty, Hendren, and Katz 2016; Chyn 2018; Chetty and Hendren 2018; Nakamura, Sigurdsson, and Steinsson 2021; Deutscher 2020) and long-run schooling (Chetty, Hendren, and Katz 2016; Chyn 2018; Chetty and Hendren 2018; Laliberté 2021). In terms of adverse outcomes, the effects of moving at early ages to higher opportunity neighborhoods is negative for teenage pregnancy (Chetty and Hendren 2018), incarcerations (Chetty et al. 2020c), and hospitalizations (Pollack et al. 2019). Thus, increased childhood exposure to better neighborhood environments generates beneficial impacts on long-run economic, schooling, social, and health outcomes.

Mechanisms for Children

This section discusses the evidence on several mechanisms thought to mediate the impacts of childhood neighborhoods on long-run outcomes: school quality, peer influences, pollution, exposure to violence, and criminal justice policies. Each of these channels varies substantially across US cities and neighborhoods due to socioeconomic and race-based residential segregation. Recent studies provide compelling causal evidence in support of each of these mechanisms. In contrast, the recent evidence suggests the causal component of childhood neighborhood exposure effects is *not* mediated by improvements in parental income.

First, lower poverty neighborhoods might also have greater school resources and better-performing schools. During the past decade or so, many studies have used lottery-based admissions to estimate, using experimental methods, the effect of attending schools with higher levels of teacher quality or school value-added. These studies have found that attending higher-quality schools improves academic achievement (as measured by standardized exams) increases postsecondary education attendance, and reduces the incidence of risky behaviors such as criminal

[10] In contrast, Oreopoulos (2003) is a compelling quasi-experimental study that finds no evidence of childhood neighborhoods affecting adult economic outcomes. Specifically, he compares long-run outcomes for children who were assigned to live in different public housing projects in Toronto and finds that the children's long-run labor market outcomes are not systematically related to the neighborhood environments surrounding their residential public housing projects. One possible explanation is that neighborhood effects operate at a hyperlocal level (limited to the area within the public housing project), and the environment within public housing projects is similar across projects despite substantial variation in the broader surrounding neighborhood environments.

activity or teenage pregnancy (Hastings and Weinstein 2008; Deming 2011; Dobbie and Fryer 2015). In addition, research using quasi-experimental methods has shown that increases in school funding have large impacts on academic achievement, educational attainment, and adult earnings and health outcomes for students from low-income families (Jackson, Johnson, and Persico 2016; Jackson and Mackevicius 2021). In this vein, Laliberté (2021) finds that 50 percent or more of the gains in educational attainment from moving to a better neighborhood in Montreal is driven by increased access to higher-quality schools.

However, the findings from the Moving to Opportunity experiment and the Chicago public housing demolitions indicate that increased academic achievement as measured by standardized test scores was not a mediating factor that led to the realized increases in college attendance and improved adult incomes for younger children moving to lower-poverty areas (Ludwig et al. 2013; Jacob 2004; Chyn 2018). One possibility is that schools in more affluent neighborhoods may still play a role for less advantaged students through improvements in non-cognitive skills and non-academic schooling outcomes such as disciplinary infractions. Jackson (2018) finds strong evidence that teacher value-added in non-cognitive skills as measured by non-test score behaviors (absences, suspensions, course grades, and grade repetition) is distinct from test score effects and has substantial impacts on longer-run student outcomes. Childhood exposure effects from schools and broader neighborhood environments could also partially be mediated by factors associated with cultural adaptability, as suggested by the finding that moves to lower-poverty areas for younger Black children in Moving to Opportunity increased the use of Standard American English as compared to African-American Vernacular English (Rickford et al. 2015).

Second, strong evidence on the impact of the influence of peers in one's community comes from studies of criminal behavior. Damm and Dustmann (2014) study a Danish natural experiment generated by a policy that quasi-randomly assigned refugee households to municipalities throughout the country: refugee children assigned to areas with higher shares of youth criminals are significantly more likely to have later-life criminal convictions. Billings, Deming, and Ross (2019) study children within small neighborhood areas in North Carolina and show that their likelihood of being arrested together (that is, being criminal partners) is higher when they attend the same school, especially for neighborhood peers who are the same race and gender.

Third, high-poverty neighborhoods typically have the greatest exposure to air pollution, water quality problems, and lead (Bernard and McGeehin 2003; Colmer et al. 2020). For children, greater exposure to air pollution has negative impacts on early-life health (Chay and Greenstone 2003; Currie and Walker 2011), human capital (Heissel, Persico, and Simon forthcoming), and labor market outcomes (Isen, Rossin-Slater, and Walker 2017). Childhood exposure to lead has negative effects on a wide range of outcomes (Aizer et al. 2018; Aizer and Currie 2019). High-poverty and high-minority share neighborhoods (especially historically redlined areas) also have land surface features leading to higher temperatures and more extreme heat exposure for residents than more advantaged (and non-redlined) neighborhoods (Hoffman, Shandas, and Pendleton 2020). A growing literature indicates adverse impacts on children's learning and school performance from cumulative heat exposure (Park et al. 2020).

Fourth, exposure to neighborhood violence affects children. Sharkey (2010) studies the impact of local homicides using a difference-in-difference approach that relies on variation in the location and timing of homicides. He finds notable short-run effects for minority children: African-American children recently exposed to homicides in their block group have lower scores on vocabulary and reading assessments. Ang (2021) uses a similar research design and shows hyperlocal exposure to violence in the form of killings by police during adolescence has longer-lasting impacts in the form of reduced rates of high school graduation and college enrollment.

Fifth, local public goods related to the criminal justice system have important impacts on long-run child outcomes. Derenoncourt (2020) finds that Northern cities that received more Black migrants during the twentieth century had lower rates of upward mobility for Black children born in the 1980s. As potential meditators driving this negative impact on mobility, she shows that Black migration also resulted in greater spending on police and higher rates of incarceration. Similarly, Baran, Chyn, and Stuart (2020) provide complementary evidence suggesting that increases in county-level incarceration rates reduced rates of Black economic opportunity between 1940 and 1990. More aggressive local policing behavior and increased incarceration risk could negatively affect children by reducing the incentive to invest in human capital (Lochner 2004).

Finally, an important implication of recent housing mobility research is that causal childhood exposure effects from moving to higher opportunity areas are not driven by parental income gains. Studies of the Moving to Opportunity demonstration and Chicago public housing demolitions found no evidence that relocating to less distressed areas had impacts on the economic outcomes of adults, but both settings revealed large long-run gains for younger children (Ludwig et al. 2013; Chetty, Hendren, and Katz 2016; Chyn 2018).

Some Implications for the Design of Housing Voucher Policies

Policymakers have become increasingly concerned about the effects of neighborhoods, given persistently high and rising levels of residential segregation by income—particularly for households with children (Owens 2016; Reardon et al. 2018). Rental vouchers and housing allowance programs for low-income households are major forms of assistance provided in developed countries (Priemus, Kemp, and Varady 2005). The Housing Choice Voucher program (previously known as the Section 8 program) is the largest form of housing aid for US disadvantaged households and aids approximately 2.3 million low-income families annually (Collinson, Ellen, and Ludwig 2015). There are similarly large housing subsidy programs in the United Kingdom and Chile.

The latest generation of neighborhood-effects studies suggest three lessons for housing voucher policies that provide portable rental support to low-income families. First, designing vouchers so that families are encouraged and helped to move to low-poverty or otherwise more advantaged areas is a crucial program feature. US housing voucher recipients typically do not use their vouchers to move to high-opportunity areas (Collinson and Ganong 2018). Evidence from the Moving to Opportunity demonstration, the Baltimore Regional Housing Program (DeLuca and Rosenblatt 2017), and the more recent Creating Moves to Opportunity program in Seattle (Bergman et al. 2020) indicates that housing mobility counseling services that provide customized assistance and encouragement can notably increase the rate at which voucher recipients move to higher-opportunity areas.[11]

Second, the social benefit–cost ratio (or the marginal value of public funds) for housing voucher programs are likely highest if the vouchers are targeted to families with young children (Hendren and Sprung-Keyser 2020). Children who move to higher-opportunity areas at younger ages have longer potential childhood exposure, which consistently leads to improved long-run outcomes. This implies that the common use of voucher waitlists—where eligible families may wait years while their children age—may be ineffective relative to prioritizing families with younger children.

What are the estimated benefits and costs to targeting housing vouchers to families with young children and encouraging moves to better neighborhoods? Chetty, Hendren, and Katz (2016) provide an assessment of this question from the perspective of the public housing families that participated in the Moving to Opportunity experiment. They find that the experimental vouchers increased annual earnings by $3,477 for children whose families moved before they were age 13. Using relatively conservative assumptions, they estimate that using a voucher to move to a high-opportunity neighborhood for a typical public housing family with two young children would increase the children's lifetime earnings by $198,000 and tax revenue by $22,400 (in present value). On the cost side, their estimate is based on the cost of providing housing mobility counseling because the fiscal costs of housing vouchers are equivalent to or less than those of providing place-based public housing (Olsen 2000). In the Moving to Opportunity program, the average cost of mobility counseling was $3,789 per family who took up a voucher (Goering et al. 1999). The findings suggest that the benefits substantially exceed the cost of providing a targeted voucher with mobility counseling instead of traditional public

[11] Recent studies show that several alternatives to intensive mobility counseling (such as financial incentives or low-intensity counseling) are much less effective in increasing the share of voucher holders moving to high-opportunity neighborhoods (Bergman et al. 2020; Schwartz, Mihaly, and Gala 2017). For example, one prominent approach, termed Small Area Fair Market Rents (SAFMRs), is to encourage families to move to higher-opportunity areas by offering higher voucher payment standards in high-rent ZIP codes. Collinson and Ganong (2018) and Bergman et al. (2020) find that SAFMRs induced modest increases in moves to higher-quality neighborhoods in Dallas and in Seattle-King County, but not nearly as large an impact as when SAFMRs are combined with customized mobility counseling as in the Creating Moves to Opportunity demonstration.

housing support, both for taxpayers and for the low-income families with young children themselves.

Finally, a third lesson is that using vouchers to facilitate low-income households to move to higher-opportunity neighborhoods within the same metropolitan area is unlikely on its own to improve the economic outcomes of adults in the relocating households. Macroeconomic or regional policies that increase overall local economic activity and labor market tightness appear more promising. Much research has found that economic conditions at the commuting zone (or state) level have strong impacts on the employment outcomes of minorities and those with less education (Hoynes, Miller, and Schaller 2012). Recent literature also uncovers persistent effects on employment outcomes from commuting zone-level recession shocks (Yagan 2019), declines in manufacturing (Charles, Hurst, and Schwartz 2019), and international trade shocks (Autor, Dorn, and Hanson 2013). Policies that improve the human capital, occupational skills, and connections to employers of low-wage workers living in high-poverty areas also have potential. An experimental evaluation of the Jobs Plus program, an employment program operating in high-poverty public housing projects, found long-lasting positive impacts on earnings (Riccio 2010).[12] In addition, some combination of better access to public transportation and housing opportunities could lower job search and commuting costs, thereby improving work outcomes for residents of disadvantaged neighborhoods. Holzer, Quigley, and Raphael (2003) found such a pattern in the case of the expansion of the Bay Area Rapid Transit (BART) system in the San Francisco Bay Area.

Discussion and Conclusion

In the past two decades, new experimental and quasi-experimental studies have pushed the frontier of research on neighborhood effects. The work surveyed indicates that residential neighborhoods within a metropolitan area matter for adult health and well-being but have little causal impact on contemporaneous adult labor market outcomes (at least for the heads of low-income households). Adult economic outcomes are shaped more by overall commuting zone or regional labor market opportunities. In contrast, the emerging consensus for children is that living in a higher-opportunity neighborhood has substantial beneficial causal impacts on a number of socioeconomic outcomes.

How do the results from the recent studies discussed in this essay change the interpretation of previous studies of neighborhood effects? There are two main implications. First, the findings strongly imply that traditional observational studies of the neighborhood effects are likely to suffer from substantial selection bias leading to overestimates of neighborhood influences on adult economic outcomes within a

[12] Sectoral employment programs also appear particularly promising for disadvantaged workers. Evidence from the WorkAdvance demonstration and the Year Up program show that providing training and placement into higher-wage occupations notably improves worker outcomes (Katz et al. 2020).

metropolitan area. For example, several studies have used non-experimental methods to study adult movers and found large effects on labor market outcomes (Fauth, Leventhal, and Brooks-Gunn 2004; Weinberg, Reagan, and Yankow 2004; Clampet-Lundquist and Massey 2008). Although selection bias is not the only explanation for the discrepancy between earlier observational studies and more recent work based on experiments and plausible quasi-experiments, Harding et al. (2021) provide evidence suggesting that selection bias plays a large role in driving the findings of traditional non-experimental studies of neighborhood effects on adult economic outcomes.

Second, recent findings reshape our understanding of the nature of neighborhood effects for children. Specifically, the analyses by Chetty and Hendren (2018), Chetty, Hendren, and Katz (2016), and Chyn (2018) provide strong evidence that neighborhoods affect outcomes through childhood exposure effects. The Moving to Opportunity experiment generated beneficial impacts on long-run economic outcomes of moves to higher-opportunity areas only for younger children who received a larger "dosage" of childhood exposure to improved neighborhood environments relative to their older counterparts. Disruption costs of moves across different types of neighborhoods could potentially outweigh the small exposure gains for older children.

We conclude by discussing several directions for further research. Future work related to mechanisms remains an ongoing research issue. For example, we know little about the relative importance of the different mechanisms that are typically "bundled" together within a neighborhood. In other words, how much does school quality matter relative to other characteristics of a local area such as peers or neighborhood safety? A better understanding of the weight of each of these neighborhood factors may improve policy responses. The deep integration of qualitative (ethnographic) research into experimental and quasi-experimental research designs, as in the Moving to Opportunity and the Creating Moves to Opportunity projects, also represents a promising direction to generate more nuanced and realistic insights into the mechanisms behind neighborhood effects (Clampet-Lundquist et al. 2011; Bergman et al. 2020).

More work is also needed to understand both the general equilibrium effects of scaling-up housing mobility policies and the impact of other policies that shape residential choice. For example, increases in the share of low-income to high-income residents in high-opportunity neighborhoods could generate aggregate gains, because neighborhoods appear to matter less for outcomes of high-income children (Chetty et al. 2020a) and may be desirable on distributional grounds. Changes in the supply of housing might also occur in the long-run from shifts in housing demand across neighborhoods due to housing mobility programs. These general equilibrium effects could be quantified through research methods combining experimental and quasi-experimental sources of variation in neighborhood choices with more structural approaches as in Galiani, Murphy, and Pantano (2015), Davis, Gregory, and Hartley (2019), Diamond and McQuade (2019), and Chyn and Daruich (2021). In addition, understanding the effects of local land-use regulations (like restrictions on multi-family housing) and housing market discrimination on access to high-opportunity neighborhoods for low-income and minority families remains a crucial related research area (Glaeser 2017; Rothstein 2017).

A final frontier research area involves the estimation of the impact of place-based policies to improve low-income neighborhoods on the intended beneficiaries—the incumbent (preexisting) adult residents and their children. Place-based policies focusing on business tax incentives such as Enterprise Zones and Opportunity Zones do not appear to be effective in improving job creation and economic opportunities in low-income areas (Bartik 2020). Public housing redevelopment efforts via the federal HOPE VI program have improved the trajectories of high-poverty and racially segregated neighborhoods—but possibly by displacing poorer and non-white residents (Tach and Emory 2017). More comprehensive and community-driven, place-based investment policies such as the federal Empowerment Zones appear to have improved area economic outcomes in repeated cross-section analyses (Busso, Gregory, and Kline 2013), but it is less clear if the gains accrue to preexisting residents or reflect changes in neighborhood sorting and accrue to in-migrants.

Newly available longitudinal administrative data sets should allow future research to examine effects of place-based policies on preexisting residents. Haltiwanger et al. (2020) is a start in this direction, specifically for understanding the impacts of place-based urban renewal programs such as the HOPE VI public housing demolitions. Similarly, Garin and Rothbaum (2020) link individuals from the 1940 Census of Population to their economic outcomes many years later in Social Security earnings data. They exploit quasi-experimental variation across counties in the construction of government-financed manufacturing plants for World War II, finding substantial positive impacts on local economic development and persistent gains in the adult earnings of children who lived in treated counties prior to the war. Further analyses of contemporary place-based policies and community development programs using linked administrative data sets (from the US Census Bureau and for many European countries) would be a valuable addition.

■ *We thank Peter Bergman, Raj Chetty, Nathan Hendren, Jens Ludwig, and Christopher Palmer for helpful suggestions and comments. We also thank Camille Baker and the team at Opportunity Insights for creating the maps that are included in this work.*

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5): 1832–56.

Agan, Amanda Y., and Sonja B. Starr. 2020. "Employer Neighborhoods and Racial Discrimination." NBER Working Paper 28153.

Aizer, Anna, and Janet Currie. 2019. "Lead and Juvenile Delinquency: New Evidence from Linked Birth, School, and Juvenile Detention Records." *Review of Economics and Statistics* 101(4): 575–87.

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2018. "Do Low Levels of Blood Lead Reduce Children's Future Test Scores?" *American Economic Journal: Applied Economics* 10 (1): 307–41.

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development before Age Five." In *Handbook of Labor Economics*, Vol. 4B, edited by David Card and Orley Ashenfelter, 1315–1486. Amsterdam: Elsevier.

Ang, Desmond. 2021. "The Effects of Police Violence on Inner-City Students." *Quarterly Journal of Economics* 136 (1): 115–68.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

Baran, Cavit, Eric T. Chyn, and Bryan A. Stuart. 2020. "The Great Migration and Educational Opportunity." Unpublished.

Barnhardt, Sharon, Erica Field, and Rohini Pande. 2017. "Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India." *American Economic Journal: Applied Economics* 9 (1): 1–32.

Bartik, Timothy J. 2020 "Using Place-Based Jobs Policies to Help Distressed Communities." *Journal of Economic Perspectives* 34 (3): 99–127.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F. Katz, and Christopher Palmer. 2020. "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice." https://opportunityinsights.org/wp-content/uploads/2019/08/cmto_paper.pdf.

Bernard, S.M., and M.A. McGeehin. 2003. "Prevalence of Blood Lead Levels >=5 g/dL among US Children 1 to 5 Years of Age and Socioeconomic and Demographic Factors Associated with Blood of Lead Levels 5 to 10 g/DL, Third National Health and Nutrition Examination Survey, 1988–1994." *Pediatrics* 112 (6): 1308–13.

Billings, Stephen B., David J. Deming, and Stephen L. Ross. 2019. "Partners in Crime." *American Economic Journal: Applied Economics* 11 (1): 126–50.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review* 105 (2): 477–503.

Boustan, Leah Platt. 2016. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton: Princeton University Press.

Brock, William A., and Steven N. Durlauf. 2001. "Discrete Choice with Social Interactions." *Review of Economic Studies* 68 (2): 235–60.

Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Kato Klebanov, and Naomi Sealand. 1993. "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology* 99 (2): 353–95.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence from a Field Experiment on Financial Decisions." *Econometrica* 82 (4): 1273–1301.

Busso, Matias, Jesse Gregory, and Patrick Kline. 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review* 103 (2): 897–947.

Charles, Kerwin, Erik Hurst, and Mariel Schwartz. 2019. "The Transformation of Manufacturing and the Decline in US Employment." *NBER Macroeconomics Annual* 33.

Chay, Kenneth Y., and Michael Greenstone. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118 (3): 1121–67.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020a. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." Opportunity Insights, Harvard University, January.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020b. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility: Dataset." https://www.opportunityatlas.org/ (accessed November 11, 2019).

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020c. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *Quarterly Journal of Economics* 135 (2): 711–83.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133 (3): 1107–62.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016a. "The Association Between Income and Life Expectancy in the United States, 2001–2014." *JAMA* 315 (16): 1750–66.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016b. "The Health Inequality Project: Dataset" https://www.health-inequality.org/data/ (accessed January 17, 2020).

Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *The American Economic Review* 108 (10): 3028–56.

Chyn, Eric, and Diego Daruich. 2021. "The Equilibrium Effects of Neighborhood Interventions." Unpublished.

Chyn, Eric, and Kareem Haggag. 2020. "Moved to Vote: The Long-Run Effects of Neighborhoods on Political Participation." NBER Working Paper 26515.

Chyn, Eric, and Lawrence Katz. 2020. "Neighborhoods Matter: Assessing the Evidence for Place Effects: Dataset." American Economic Association [publisher]. Inter-university Consortium for Political and Social Research [distributor].

Clampet-Lundquist, Susan, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2011. "Moving Teenagers out of High-Risk Neighborhoods: How Girls Fare Better than Boys." *American Journal of Sociology* 116 (4): 1154–89.

Clampet-Lundquist, Susan, and Douglas S. Massey. 2008. "Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment." *American Journal of Sociology* 114 (1): 107–43.

Clemens, Michael A. 2013. "Why Do Programmers Earn More in Houston than Hyderabad? Evidence from Randomized Processing of US Visas." *American Economic Review* 103 (3): 198–202.

Collins, William J., and Marianne H. Wanamaker. 2014. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics* 6 (1): 220–52.

Collinson, Robert, Ingrid Gould Ellen, and Jens Ludwig. 2015. "Low-Income Housing Policy." In *Economics of Means-Tested Transfer Programs in the United State*s, Vol. 2, edited by Robert A. Moffitt, 59–126. Chicago: University of Chicago Press.

Collinson, Robert, and Peter Ganong. 2018. "How Do Changes in Housing Voucher Design Affect Rent and Neighborhood Quality?" *American Economic Journal: Economic Policy* 10 (2): 62–89.

Colmer, Jonathan, Ian Hardman, Jay Shimshack, and John Voorheis. 2020. "Disparities in $PM_{2.5}$ Air Pollution in the United States." *Science* 369 (6503): 575–78.

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3 (1): 65–90.

Damm, Anna Piil, and Christian Dustmann. 2014. "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?" *American Economic Review* 104 (6): 1806–32.

Davis, Morris A., Jesse Gregory, and Daniel A. Hartley. 2019. "The Long Run Effects of Low-Income Housing on Neighborhood Composition." Unpublished.

DeLuca, Stefanie, Greg J. Duncan, Mierce Keels, and Ruby M. Mendenhall. 2010. "Gautreaux Mothers and Their Children: An Update." *Housing Policy Debate* 20 (1): 7–25.

DeLuca, Stefanie, and Peter Rosenblatt. 2017. "Walking Away from The Wire: Housing Mobility and Neighborhood Opportunity in Baltimore." *Housing Policy Debate* 27 (4): 519–46.

Deming, David J. 2011. "Better Schools, Less Crime?" *Quarterly Journal of Economics* 126 (4): 2063–2115.

Derenoncourt, Ellora. 2020. "Can You Move to Opportunity? Evidence from the Great Migration." Unpublished.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." *American Economic Journal: Applied Economics* 10 (2): 202–33.

Deryugina, Tatyana, and David Molitor. 2020. "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina." *American Economic Review* 110 (11): 3602–33.

Deutscher, Nathan. 2020. "Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia." *American Economic Journal: Applied Economics* 12 (2): 220–49.

Diamond, Rebecca, and Timothy J. McQuade. 2019. "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development." *Journal of Political Economy* 127 (3): 1063–1117.

Dobbie, Will J., and Roland G. Fryer. 2015. "The Medium-Term Impacts of High-Achieving Charter Schools." *Journal of Political Economy* 123 (5): 985–1037.

Duflo, Esther and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics* 118 (3): 815–42.

Durlauf, Steven N. 2004. "Neighborhood Effects." In *Handbook of Regional and Urban Economic*s, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2173–2242. Amsterdam: Elsevier.

Fahle, Erin M., Benjamin R. Shear, Dementra Kalogrides, Sean F. Reardon, Belen Chavez, and Andrew D. Ho. 2019. "Stanford Education Data Archive: Technical Documentation (Version 3.0)." Stanford Digital Repository. http://purl.stanford.edu/db586ns4974. (accessed December 12, 2019).

Fauth, Rebecca C., Tama Leventhal, and Jeanne Brooks-Gunn. 2004. "Short-Term Effects of Moving from Public Housing in Poor to Middle-Class Neighborhoods on Low-Income, Minority Adults' Outcomes." *Social Science & Medicine* 59 (11): 2271–84.

Galiani, Sebastian, Alvin Murphy, and Juan Pantano. 2015. "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment." *American Economic Review* 105 (11): 3385–3415.

Galster, George, and Patrick Sharkey. 2017. "Spatial Foundations of Inequality: A Conceptual Model and Empirical Overview." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3 (2): 1–33.

Garin, Andrew, and Jonathan Rothbaum. 2020. "Was the Arsenal of Democracy an Engine of Mobility? Public Investment and the Roots of Mid-century Manufacturing Opportunity." Unpublished.

Glaeser, Edward. 2017. "Reforming Land Use Regulations." *The Brookings Institution*, April 24. https://www.brookings.edu/research/reforming-land-use-regulations/.

Glaeser, Edward L., David Laibson, and Bruce Sacerdote. 2002. "An Economic Approach to Social Capital." *Economic Journal* 112 (483): 437–458.

Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics* 111 (2): 507–48.

Goering, John, Joan Kraft, Feins Judith, Dennis McInnis, Mary Holin, and Huda Elhassan. 1999. "Moving to Opportunity for Fair Housing Demonstration Program: Current Status and Initial Findings." https://www.huduser.gov/publications/pdf/mto.pdf.

Graham, Bryan S. 2018. "Identifying and Estimating Neighborhood Effects." *Journal of Economic Literature* 56 (2): 450–500.

Haltiwanger, John C., Mark J. Kutzbach, Giordano E. Palloni, Henry Pollakowski, Matthew Staiger, and Daniel Weinberg. 2020. "The Children of HOPE VI Demolitions: National Evidence on Labor Market Outcomes." NBER Working Paper 28157.

Harding, David, Lisa Sanbonmatsu, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Matthew Sciandra, and Jens Ludwig. 2021. "Evaluating Contradictory Experimental and Non-Experimental Estimates of Neighborhood Effects on Adult Economic Outcomes." NBER Working Paper 28454.

Hastings, Justine, and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–1414.

Heckman, James J., and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6: 689–733.

Heissel, Jennifer A., Claudia Persico, and David Simon. Forthcoming. "Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance." *Journal of Human Resources***.**

Hendren, Nathaniel, and Benjamin Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics* 135 (3): 1209–1318.

Hoffman, Jeremy, Vivek Shandas, and Nicholas Pendleton. 2020. "The Effects of Historical Housing Policies on Resident Exposure to Intra-Urban Heat: A Study of 108 US Urban Areas." *Climate* 8 (1).

Holzer, Harry J. 1991. "The Spatial Mismatch Hypothesis: What Has the Evidence Shown?" *Urban Studies* 28 (1):105–22.

Holzer, Harry J., John M Quigley, and Steven Raphael. 2003. "Public Transit and the Spatial Distribution of Minority Employment: Evidence from a Natural Experiment." *Journal of Policy Analysis and Management* 22 (3): 415–41.

Hoynes, Hilary, Douglas L. Miller, and Jessamyn Schaller. 2012. "Who Suffers during Recessions?" *Journal of Economic Perspectives* 26 (3): 27–48.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. 2017. "Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970." *Journal of Political Economy* 125 (3): 848–902.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy* 126 (5): 2072–2107.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on

Educational and Economic Outcomes: Evidence from School Finance Reforms." *Quarterly Journal of Economics* 131 (1): 157–218.

Jackson, C. Kirabo, and Clare Mackevicius. 2021. "The Distribution of School Spending Impacts." NBER Working Paper 28517.

Jacob, Brian A. 2004. "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94 (1): 233–58.

Jencks, Christopher, and Susan E. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner-City Poverty in the United States*, edited by Laurence Lynn Jr., and Michael G.H. McGeary. Washington, D.C: National Academy Press.

Johnson, Janna E., and Evan J. Taylor. 2019. "The Long Run Health Consequences of Rural-urban Migration." *Quantitative Economics* 10 (2): 565–606.

Kain, John F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *Quarterly Journal of Economics* 82 (2): 175–97.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116 (2): 607–54.

Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg. 2020. "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance." NBER Working Paper 28248.

Kaufman, Julie E., and James Rosenbaum. 1992. "The Education and Employment of Low-Income Black Youth in White Suburbs." *Educational Evaluation and Policy Analysis* 14 (3): 229–40.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

Laliberté, Jean-William. 2021. "Long-Term Contextual Effects in Education: Schools and Neighborhoods." *American Economic Journal: Economic Policy* 13 (2): 336–77.

Lochner, Lance. 2004. "Education, Work, and Crime: A Human Capital Approach." *International Economic Review* 45 (3): 811–43.

Loury, Glenn C. 1977. "A Dynamic Theory of Racial Income Differences." In *Women, Minorities, and Employment Discrimination*, edited by Phyllis A. Wallace and Annette LaMond, 153–86. Lexington, MA: Lexington Books.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science* 337 (6101): 1505–10.

———. 2013. "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review* 103 (3): 226–31.

Ludwig, Jens, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade. 2011. "Neighborhoods, Obesity, and Diabetes—A Randomized Social Experiment." *New England Journal of Medicine* 365 (16): 1509–19.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.

McKenzie, David, Steven Stillman, and John Gibson. 2010. "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration." *Journal of the European Economic Association* 8 (4): 913–45.

Mendenhall, Ruby, Stefanie DeLuca, and Greg Duncan. 2006. "Neighborhood Resources, Racial Segregation, and Economic Mobility: Results from the Gautreaux Program." *Social Science Research* 35 (4): 892–923.

Miller, Sarah, and Cindy K. Soo. 2021. "Do Neighborhoods Affect the Credit Market Decisions of Low-Income Borrowers? Evidence from the Moving to Opportunity Experiment." *Review of Financial Studies*. 34 (2): 827–63.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. 2021. "The Gift of Moving: Intergenerational Consequences of a Mobility Shock." https://eml.berkeley.edu/~enakamura/papers/giftofmoving.pdf.

Olsen, Edgar O. 2000. "The Cost-Effectiveness of Alternative Methods of Delivering Housing Subsidies." University of Virginia Working Paper 351.

Oreopoulos, Philip. 2003. "The Long-Run Consequences of Living in a Poor Neighborhood." *Quarterly Journal of Economics* 118 (4): 1533–75.

Owens, Ann. 2016. "Inequality in Children's Contexts: Income Segregation of Households with and without Children." *American Sociological Review* 81 (3): 549–74.

Park, R. Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith. 2020. "Heat and

Learning." *American Economic Journal: Economic Policy* 12 (2): 306–39.

Pollack, Craig Evan, Amanda L. Blackford, Shawn Du, Stefanie Deluca, Rachel L.J. Thornton, and Bradley Herring. 2019. "Association of Receipt of a Housing Voucher with Subsequent Hospital Utilization and Spending." *JAMA* 322 (21): 2115–24.

Popkin, Susan J., James E. Rosenbaum, and Patricia M. Meaden. 1993. "Labor Market Experiences of Low-Income Black Women in Middle-Class Suburbs: Evidence from a Survey of Gautreaux Program Participants." *Journal of Policy Analysis and Management* 12 (3): 556–73.

Priemus, Hugo, Peter A. Kemp, and David P. Varady. 2005. "Housing Vouchers in the United States, Great Britain, and the Netherlands: Current Issues and Future Perspectives." *Housing Policy Debate* 16 (3–4): 575–609.

Reardon, Sean F., Kendra Bischoff, Ann Owens, and Joseph B. Townsend. 2018. "Has Income Segregation Really Increased? Bias and Bias Correction in Sample-Based Segregation Estimates." *Demography* 55 (6): 2129–60.

Riccio, James. 2010. *Sustained Earnings Gains for Residents in a Public Housing Jobs Program: Seven-Year Findings from the Jobs-Plus Demonstration*. MDRC: New York City.

Rickford, John R., Greg J. Duncan, Lisa A. Gennetian, Ray Yun Gou, Rebecca Greene, Lawrence F. Katz, Ronald C. Kessler et al. 2015. "Neighborhood Effects on Use of African-American Vernacular English." *PNAS* 112 (38): 11817–22.

Rosenbaum, James E., 1995. "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program." *Housing Policy Debate* 6 (1): 231–69.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: Liveright Publishing.

Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL.: University of Chicago Press.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research." *Annual Review of Sociology* 28: 443–78.

Sampson, Robert J., and Patrick Sharkey. 2008. "Neighborhood Selection and the Social Reproduction of Concentrated Racial Inequality." *Demography* 45: 1–29.

Schwartz, Heather L., Kata Mihaly, and Breann Gala. 2017. "Encouraging Residential Moves to Opportunity Neighborhoods: An Experiment Testing Incentives Offered to Housing Voucher Recipients." *Housing Policy Debate* 27 (2): 230–60.

Sharkey, Patrick. 2010. "The Acute Effect of Local Homicides on Children's Cognitive Performance." *PNAS* 107 (26): 11733–38.

Sharkey, Patrick, and Jacob W. Faber. 2014. "Where, When, Why, and for Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects." *Annual Review of Sociology* 40: 559–79.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card, Volume 3, 1761–1800. Amsterdam: Elsevier.

Tach, Laura, and Allison Dwyer Emory. 2017. "Public Housing Redevelopment, Neighborhood Change, and the Restructuring of Urban Inequality." *American Journal of Sociology* 123 (3): 686–739.

Topa, Giorgio, and Yves Zenou. 2015. "Neighborhood and Network Effects." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 561–624. Amsterdam: Elsevier.

Votruba, Mark Edward, and Jeffrey R. Kling. 2009. "Effects of Neighborhood Characteristics on the Mortality of Black Male Youth: Evidence from Gautreaux, Chicago." *Social Science & Medicine* 68 (5): 814–23.

Weinberg, Bruce A., Patricia B. Reagan, and Jeffrey J. Yankow. 2004. "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data." *Journal of Labor Economics* 22 (4): 891–924.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: University of Chicago Press.

Wodtke, Geoffrey T., David J. Harding, and Felix Elwert. 2011. "Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation." *American Sociological Review* 76 (5): 713–36.

Wood, David, Neal Halfon, Debra Scarlata, Paul Newacheck, and Sharon Nessim. 1993. "Impact of Family Relocation on Children's Growth, Development, School Function, and Behavior." *Journal of the American Medical Association* 270 (11): 1334–38.

Yagan, Danny. 2019. "Employment Hysteresis from the Great Recession." *Journal of Political Economy* 127 (5): 2505–58.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 223–248

College Majors, Occupations, and the Gender Wage Gap

Carolyn M. Sloane, Erik G. Hurst, and Dan A. Black

In reckoning with the presence and evolution of gender gaps in wages and employment, scholars have identified that occupation matters. One's occupation reflects the influence of the narrowing effects of both formal barriers presented by the market and an individual's personal choices and interests. In terms of policy, it would be helpful to understand the extent to which gender gaps in wages and employment are predetermined by sorting that happens before a worker enters the labor market and lands in an occupation.

For many workers, pre-market sorting is unobservable. For college-educated workers, such sorting can be observed in one's undergraduate major. Before entering the labor market, college-educated workers are faced with a menu of potential undergraduate college majors that is nearly as varied as the opportunities to specialize in an occupation once in the market. Although there has been convergence over time, men and women continue to cluster in different undergraduate majors. In this paper, we show that generations of college-educated women in the United States have sorted into majors that systematically lower their potential wages relative to men. Even when men and women have sorted into the same major, women subsequently sort into occupations with lower potential wages.

■ *Carolyn M. Sloane is Assistant Professor of Economics, University of California, Riverside, Riverside, California. Erik G. Hurst is Frank P. and Marianne R. Diassi Distinguished Service Professor of Economics, University of Chicago Booth School of Business, Chicago, Illinois. Dan A. Black is Professor of Public Policy, University of Chicago, Harris School of Public Policy, Chicago, Illinois. Their email addresses are csloane@ucr.edu, erik.hurst@chicagobooth.edu, and danblack@uchicago.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.223.

We do not attempt to unravel the underlying dynamics that lead men and women to cluster into different majors and occupations. Instead, we document the extent to which this sorting occurs before and after entry into the market. We define a new set of summary statistics that allow the reader to consider sorting in a price space, better understand the linkages between major and occupation, and assess the independent contributions of major and occupation to gender gaps in wages and employment.

In order to do this, we take advantage of a large multi-cohort data set linking detailed undergraduate major to subsequent occupations in a way that was not available until recently. We use data from the American Community Survey (ACS) that includes questions about undergraduate major for millions of college-educated individuals. Starting in 2009, these questions were asked for people of all ages, which allows us to explore how patterns of sorting into both majors and occupations have changed across cohorts ranging from those born in the 1950s through the early 1990s. The data also include information about individuals' current labor market status, allowing us to link undergraduate major to occupation and wages for multiple cohorts of men and women.

To weigh the economic implications, we introduce the concept of a "potential wage" as the earnings one would receive, largely absent the effect of market influences such as discrimination or disruptions in job tenure. In other words, the potential wage assigns everyone within a category of occupation or major the same hourly wages—that of the median middle-aged, US-born, White male within the category. In doing so, we hope to better home in on the question: How much of the gender wage gap among college graduates can be explained by sorting into undergraduate majors with different earnings potential?

We find that women are systematically sorted into majors with lower potential wages relative to men. For example, Aerospace Engineering, one of the highest potential wage majors, is 88 percent male, while Early Childhood Education, one of the lowest potential wage majors, is 97 percent female. We also find that such patterns are long-standing and have been slow to converge. Overall, college-educated women born in the 1950s matriculated with majors that had potential wages 12 percent lower than men from their cohort. That gap fell to about 9 percent for the 1990 birth cohort. Even after some convergence in major sorting between men and women during the last 40 years, the youngest birth cohorts of women are still sorted into majors with lower potential wages than their male peers. Intriguingly, much of the convergence in major sorting between men and women occurred between the 1950 and 1975 birth cohorts, with a modest divergence for recent cohorts.

We put this data and a set of new summary statistics to use to address the linkages between pre-market and market specialization. In other words, we answer: how has occupational sorting conditional on major evolved across generations of US college graduates? We find that while women are sorted into occupations with lower potential wages conditional on major, this gap is closing somewhat over time. For the 1950 birth cohort, for example, women on average sorted to occupations with 11 percent lower potential earnings relative to otherwise similar men with the same majors. This gap narrowed to about 9 percent for the 1990 birth cohort.

Almost all of the convergence occurred within highest potential earning majors. For example, women from the 1950 cohort who majored in Engineering—a high potential earning major—sorted into occupations with potential wages that were 14 percent lower than men from the same cohort who also majored in Engineering. For the 1990 birth cohort, however, women who majored in Engineering ended up working in occupations with roughly the same potential wages as their male peers.

We close by assessing within a regression framework the separate impacts of college major and occupation on the gender wage gap among college graduates. Gender differences in major sorting explain a substantive portion of the gender wage gap among college graduates above and beyond what is explained by gender differences in occupational sorting. We also find that occupation has become a less important determinant of gender gaps in wages for US college graduates when we compare older and younger birth cohorts. In contrast, major has remained almost equally important. Overall, our results highlight the importance of understanding the social processes that lead to clustering by gender both for college major and for the occupational sorting that happens after a college major has been chosen.

Gender Differences in College Majors and Occupations

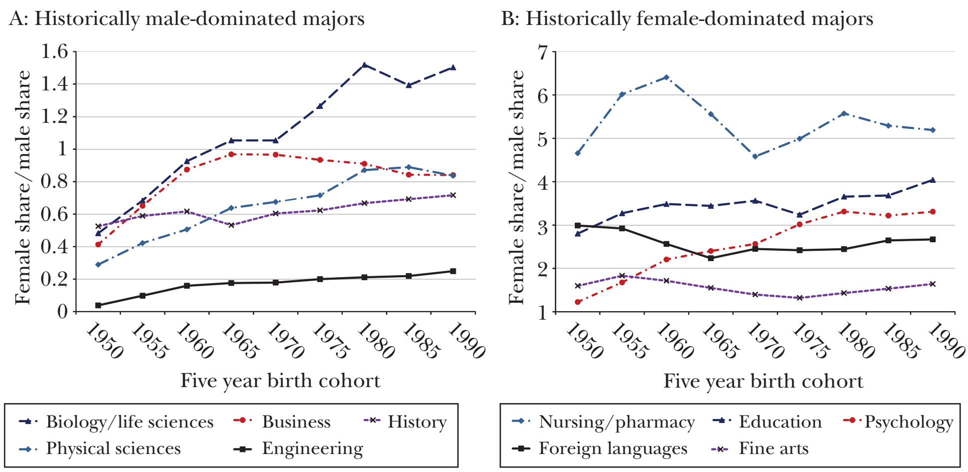
We begin by documenting the presence and evolution of gender gaps in undergraduate major sorting and how those gaps have evolved over time. For comparison, we also document trends in relative occupation for college-educated men and women. Specifically, we highlight: 1) how women from earlier birth cohorts systematically sorted into both majors and occupations that pay less than men; 2) how the gender gap in majors and occupations have narrowed for recent cohorts; and 3) how women from more recent cohorts still sort into to majors and occupations that pay less than men. To do so, we introduce a new index that measures gender differences in sorting in units of potential wage differentials.

Sorting Patterns in the Broad Categories of Majors

Our sample from the 2014–2017 American Community Survey contains information on 134 detailed majors.[1] Figure 1 graphs the ratio of females to males within broad categories of majors. For example, Engineering is a broad major category, while Civil Engineering, Chemical Engineering, and Aerospace Engineering are three of the 17 detailed majors within the broad Engineering major category. In addition, for each survey year between 2014 and 2017, we begin by assigning each

[1] Throughout the paper, we restrict our sample to those individuals who obtained a bachelor's degree. A detailed discussion of our sample restrictions and how we processed the data is found in the online Appendix available with this article at the *JEP* website.

Figure 1
Gender Differences in Selected Majors by Birth Cohort



Source: Data from the 2014–2017 ACS and are restricted to those with at least a bachelor's degree. See text and the online Appendix for additional details.
Note: These figures plot the ratio of females to males within major category. The left panel shows trends for a set of majors where men outnumber women. The right panel shows trends for a set of majors where women outnumber men.

individual a five-year birth cohort based on current age in the survey year.[2] Pooling the data into five-year birth cohorts has the effect of smoothing annual fluctuations.

Panel A focuses on majors historically dominated by men. Some of these majors have shown substantial gender convergence across birth cohorts. Consider the 1950 birth cohort: Engineering majors contained 20 men for every one woman. In the youngest birth cohorts in our sample, Engineering majors are still male-dominated, but the gap has narrowed. By the 1990 birth cohort, there were five men for every one woman in Engineering majors. These patterns are shown in the solid line in panel A. Similar convergence patterns are seen for Physical Science majors (like Chemistry, Physics, and Astronomy) and for Biology/Life Sciences majors (like Biology, Molecular Biology, Genetics, and Ecology). In fact, Biology/Life Sciences switched from being a major field dominated by men (for the 1950–1970 birth cohorts) to one dominated by women (the 1980–1990 birth cohorts). Business majors display a different pattern: women converged toward men between the 1950–1965 birth cohorts, a period when the Business major itself was expanding. Thereafter, in a period that was marked by a contraction overall in Business majors, women and men once again diverged. The History major has been male-dominated and experienced little convergence or divergence over subsequent birth cohorts.

[2] We center the five-year birth cohorts around years that are multiples of five: for example, the 1950 birth cohort includes all individuals born between 1948 and 1952.

Similar heterogeneity appears in historically female-dominated majors, shown in panel B. Education majors saw little convergence or divergence between women and men over the last 50 years. The fraction of all women majoring in Education declined substantially, but the decline among males was slightly larger. Likewise, there was little gender convergence over time in Nursing/Pharmacy majors.[3] Conversely, some gender convergence was seen in both Foreign Language and Fine Arts majors. Psychology majors were more likely to be populated by women in the 1950 cohort and became even more female-dominated by the 1990 cohort.

The broad patterns presented here are not new. For example, England and Li (2006) and Blau, Ferber, and Winkler (2014, ch. 8) use nationally representative data from the National Center on Education Statistics to document how gender differences in detailed undergraduate majors have diminished over time for a nationally representative sample of undergraduates. Black et al. (2008) document these patterns using the 1993 National Survey of College Graduates. Other studies like Dickson (2010) and Zafar (2013) have looked at gender differences in major sorting using administrative data from a few universities.

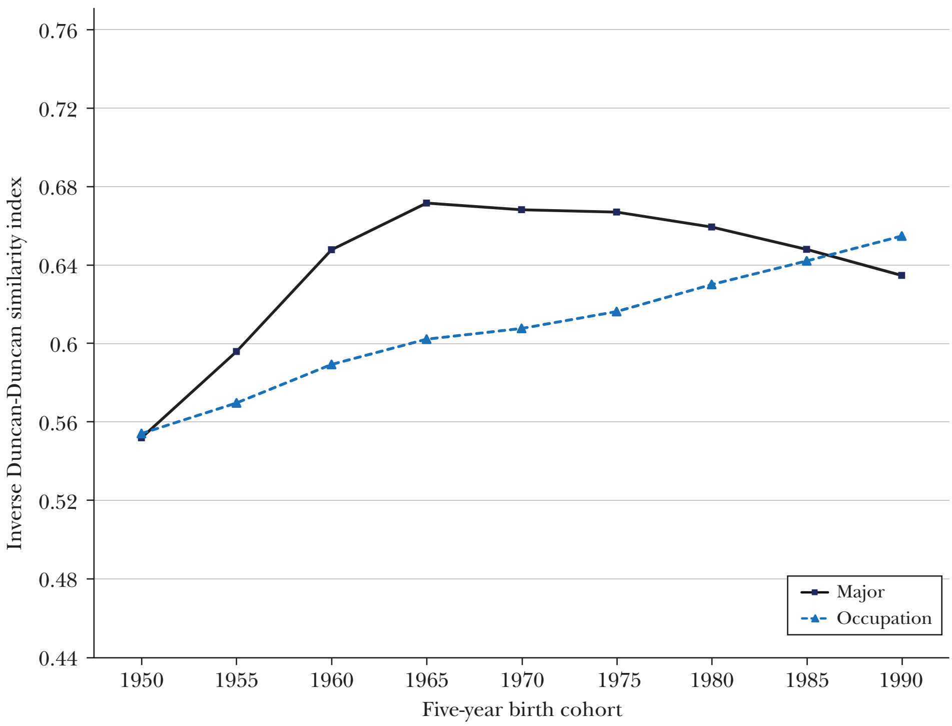
A Summary Measure of Sorting Patterns across Majors

We now use an existing method to summarize overall trends in sorting patterns across all detailed major categories. We compute an inverse Duncan-Duncan index of undergraduate major sorting patterns by gender and cohort. Separately, we compute an inverse Duncan-Duncan index of occupation sorting patterns by gender and cohort.

The classic Duncan-Duncan index for majors (or occupations) is computed by: 1) calculating the absolute within-major (or within-occupation) difference between the share of the relevant male sample in a major (or occupation) and the share of the relevant female sample in a major (or occupation); 2) summing up absolute differences over all majors (or occupations); and 3) scaling by one-half to account for the comparison of two distributions. The classic Duncan-Duncan index values range from 0 to 1. If there are no differences across majors (occupations), the index would be zero. If there is complete segregation across majors (occupations), the index would be 1.[4] To gain some intuition for this index, consider a very simple example. Say there are two majors, *A* and *B*. For men, 50 percent graduated with major *A* and 50 percent of men graduated with major *B*. For women, 70 percent graduated with major *A* and 30 percent of women graduated with major *B*. In this

[3] The broad major field referred to as "Nursing/Pharmacy" represents a combination of health-related majors: Nursing, Pharmacy, Treatment Therapy Professions, Community and Public Health, and Miscellaneous Health Medical Professions. Nursing and Pharmacy are the two largest of the combined majors.
[4] For more information on the classic Duncan-Duncan index, see Duncan and Duncan (1955).

Figure 2
Gender Similarity in Major and Occupation by Cohort



Source: Data from the 2014–2017 American Community Survey and are restricted to those with at least a bachelor's degree. See text and online Appendix for additional details.
Note: Figure plots the renormalized, inverse Duncan-Duncan indices $\left(I_c^{DD,Major}, I_c^{DD,Occ}\right)$ for different cohorts for major (solid line) and separately occupation (dashed line).

case, the Duncan-Duncan index value would equal 0.20, which corresponds to a well-integrated distribution with respect to gender.[5]

For our discussion, we renormalize these classic Duncan-Duncan indices by subtracting them from one. This step allows the reader to make easy visual comparisons with the indices introduced later in the paper: an increasing slope of any index in this paper represents a movement toward gender parity. Thus, in the simple example presented above, the inverse Duncan-Duncan index value equals 0.8 while the classic Duncan-Duncan value equals 0.2.

The solid line in Figure 2 shows the trend in the inverse Duncan-Duncan index in major sorting across different birth cohorts as computed using all 134 detailed major fields in the 2014–2017 American Community Survey. The index increased from 0.55 for the 1950 birth cohort to 0.64 for the 1990 birth cohort. There is

[5] For further discussion of this shift, see Becker, Hubbard, and Murphy (2010), Charles and Luoh (2003), DiPrete and Buchmann (2006), Goldin, Katz, and Kuziemko (2006), and Jacob (2002).

a systematic convergence in undergraduate major sorting by gender occurring between the 1950 and 1965 birth cohorts. We document a surprising new fact that there is a slight reversal in the index for recent cohorts. We also see evidence of this reversal in Figure 1 if we look at patterns among the Business and Psychology majors which experienced gender divergence in recent cohorts.

In thinking about this divergence, it is important to keep in mind a macro trend with respect to college: overall college enrollments and completions in the United States have increased, and that increase was driven by women. In our sample in the 2014–2017 American Community Survey, 30 percent of women and 34 percent of men in the 1950 birth cohort had completed at least a bachelor's degree. By the 1960 birth cohort, women had overtaken men in completion. In the 1990 birth cohort, 38 percent of women and 29 percent of men had completed at least a bachelor's degree. As more women entered college, this likely broadened in a compositional sense the set of female college graduates that naturally would be reflected in their subsequent majors. Nevertheless, the recent divergence is small relative to the convergence that occurred for older cohorts.

Sorting Patterns across Occupations

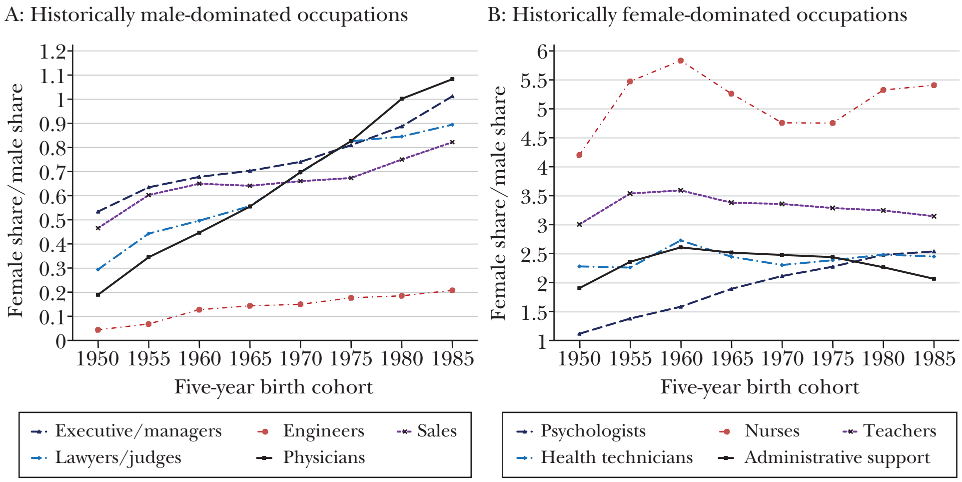
Next, we explore sorting patterns with respect to the occupations of the college graduates in our sample. In doing so, we provide a benchmark for the convergence in sorting that we see in the major space. Specifically, Figure 3 replicates the exercise from Figure 1 for a selection of male-dominated broad occupations (panel A) and female-dominated broad occupations (panel B).

We begin with male-dominated occupations as shown in panel A. College educated women from the 1950 birth cohort were 20 times less likely than men to work in the Engineering occupation. College-educated women from the 1950 birth cohort were also much less likely to work in Executive/Manager, Sales, Physician, and Lawyer occupations relative to comparable men. In all of these broad occupations, the male–female gap narrowed substantively for more recent birth cohorts.[6] By the 1985 birth cohort of college graduates, the male–female gap in the Physician and Executive/Manager broad occupations was eliminated. The gender convergence in male-dominated occupations shown in panel A of Figure 3 is consistent with the gender convergence in male-dominated majors shown in panel A of Figure 1.

We see different patterns for the female-dominated occupations in panel B of Figure 3, namely, less convergence than in the male-dominated occupations. Historically, Psychologist was a slightly more prevalent occupation for college-educated women. For recent cohorts, this occupation became increasingly more female-dominated as college women became 2.5 times more likely than men to work as Psychologists. Nurses, Health Technicians, Teachers, and Administrative Support occupations have remained female-dominated during the last 50 years.

[6] These results are consistent with the occupational sorting patterns documented in Hsieh et al. (2019).

Figure 3
Gender Differences in Selected Occupations by Birth Cohort



Source: Data from the 2014–2017 ACS and are restricted to those with at least a bachelor's degree with non-missing occupation information. See text for additional details.
Note: These figures plot the ratio of females to males within broad occupation category. The left panel shows trends for a set of occupations where men outnumber women. The right panel shows trends for a set of occupations where women outnumber men.

Thus, we do not see gender turnover from female-to-male dominance in these occupations among college graduates.

We return now to Figure 2, which moves us beyond a description of a select set of broad occupations as presented in Figure 3 to a summary of overall sorting patterns across all detailed occupation categories. For this exercise, we compute an inverse Duncan-Duncan index for occupations using the 251 distinct occupation codes reported in the 2014 to 2017 American Community Survey. As seen in Figure 2, the occupational segregation index (dashed line) is roughly similar in both level and overall trend to the major segregation index (solid line). Both indices start at a level of around 0.55 for the 1950 cohort and end at a level of around 0.65 for the 1990 cohort.

While there has a been a modest divergence in major sorting across genders for recent cohorts, convergence in occupation increased monotonically throughout.[7] The results in Figure 3 highlight that some of the convergence in the Duncan-Duncan occupation index among college graduates has been driven by relative entry of

[7] We have replicated the gender convergence in occupation using data from the historical US Censuses. This allows us to control for both cohort and age. Even conditional on age, the convergence in occupation is nearly identical to what is shown in Figure 2. See the online Appendix for additional details.

women into previously male-dominated occupations. We will explore the implications of this observation later in the paper.

The Wage Differentials of Sorting

Segregation indices such as the Duncan-Duncan index have notable shortcomings. Importantly, the Duncan-Duncan index is invariant to the earnings rank of the major field or occupation. In other words, this index tells us to what extent college-educated men and women have sorted into similar majors or occupations, but would take on the same value if all men were Physical Education majors and all women were Fine Arts majors as it would if all men were Physical Education majors and all women were Biomedical Engineering majors. These scenarios would yield vastly different earnings implications. Thus, the units of the Duncan-Duncan index do not lend themselves easily to an economic interpretation.

As an alternative measure, we develop an index that compares the impact of undergraduate major on "potential wages" by gender. In contrast to the Duncan-Duncan index, the units of the potential-wage index allow for an economic interpretation of the impact of gender differences in sorting by major. The inputs of this index also prove useful in the ensuing empirical analysis of the college gender wage gap.

A crucial input is a potential wage based on major m that is related only to differences in sorting. Specifically, we define the potential wage $\bar{Y}^{m}_{male}$ to be the median within-major labor market log wage of a group we assume faces minimal post-educational frictions in the labor market: native-born, White men between the ages of 43 and 57 with strong attachment to the labor market. For example, for anyone (male or female) who majored in Economics, we assign as their potential wage the median log wage of middle-aged, native White men who majored in Economics. In our sample, the highest potential wage majors include the broad major categories of Engineering, Mathematics and Statistics, Computer and Information Sciences, Physical Sciences, and Biology and Life Sciences. For the 1968–1977 birth cohort, women represent 16 percent, 46 percent, 24 percent, 41 percent, and 54 percent respectively of these majors in our sample in the 2014–2018 American Community Survey. The lowest potential wage majors include the broad major categories of Philosophy and Religious Studies, Education Administration and Teaching, Cosmetology Services and Culinary Arts, Fine Arts, and Public Affairs, Policy and Social Work. For the 1968–1977 birth cohort, women represent 31 percent, 77 percent, 43 percent, 58 percent, and 82 percent respectively of these majors in our sample.

This approach is intended to answer the specific question: how much would wages by gender differ based just on sorting by major, not on other market factors related to age, race, nativity, or gender. For example, this index shuts down the direct effect of mothers or caregivers experiencing more disruptions in their job tenure. Likewise, it would shut down the direct effect of discrimination with respect to promotion opportunities and other confounding factors such as potential productivity differences by gender. However, our index will necessarily absorb the fact that men and women who choose the same major may very likely end up working in different occupations; we will return to that issue later in the paper.

We formally define the potential wage index for major as:

$$I_c^{Major} = \frac{\sum_{m=1}^{M} s_{female,c}^{m} \bar{Y}_{male}^{m}}{\sum_{m=1}^{M} s_{male,c}^{m} \bar{Y}_{male}^{m}} - 1.$$

Here, I_c^{Major} is an index that measures the differential potential log wage of women of cohort c given their major m relative to the majors of men from a similar cohort. A value of $I_c^{Major} = 0$ means that the majors of women yield the same potential log wage as their male counterparts. Any deviations from zero necessarily implies that women and men sorted into different undergraduate majors. For example, a potential wage index value of –0.12 implies that women sort into majors with a potential wage that is 12 percent lower relative to males from a similar cohort.[8]

As with the inverse Duncan-Duncan index, we replicate the same exercise with respect to occupations calculating a potential wage index based on occupation, I_c^{Occ}, for college graduates. For anyone—male or female—who work in a given occupation, such as Registered Nurse, we assign as their potential wage the median log wage of older native White men who work in that occupation. Again, this index will by definition have a value of zero if men and women sort into occupations that, weighted by the numbers of men and women in each occupation, have the same potential pay. When $I_c^{Occ} < 0$, women are sorting into occupations with lower potential wages than men from a similar cohort.

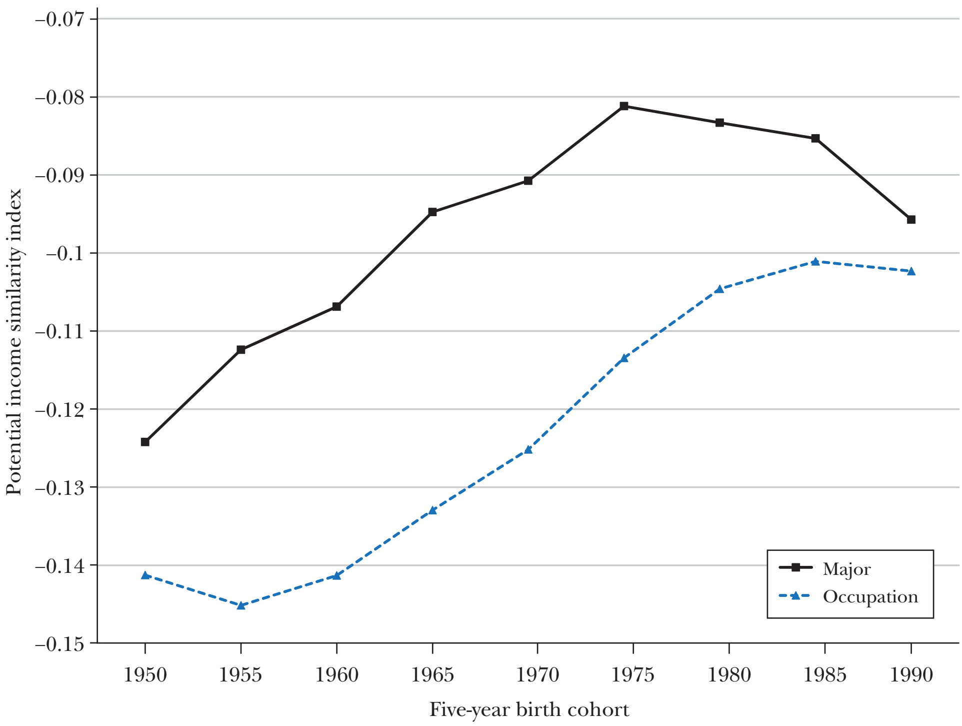
In our sample, the highest potential wage occupations include the broad occupation categories of Physicians, Lawyers and Judges, Executive, Managerial and Administrative Services, Other Technicians, and Engineers. In the 1968–1977 cohort, women account for 43 percent, 43 percent, 44 percent, 28 percent, and 14 percent respectively of the college employment in these occupations. The lowest potential wage occupations include the broad occupation categories of Housekeeping, Personal Appearance, Child Care, Food Prep and Service, and Buildings, Maintenance and Keeping. Women account for 85 percent, 81 percent, 95 percent, 58 percent, and 23 percent respectively of the college employment in these occupations in our sample.[9]

The solid line in Figure 4 shows the trend in the potential wage indices across cohorts. Notice that the vertical axis of Figure 4 includes only negative values, implying that in both major and occupation space, women have systematically sorted into fields with lower potential wages than men. As with the inverse Duncan-Duncan index, a positive slope to our potential wage indices would imply convergence between men

[8] This is similar to an index developed in Bertrand (2017). Using a similar methodology to rank undergraduate majors in the American Community Survey, Bertrand (2017) notes gender convergence at the 90th percentile, 80th percentile, and mean of log earnings across birth cohorts.

[9] We use the broad occupational categories to summarize the heterogeneity of occupations, but we calculate the index using detailed occupational codes.

Figure 4
Potential Wage Index in Major and Occupation by Cohort



Source: Data from the 2014–2017 ACS and are restricted to those with at least a bachelor's degree. See online Appendix for additional details.
Note: Figure plots the potential wage indices for major and occupation, across different cohorts of US college graduates. The solid line shows the index for major (I_c^{Major}). The dashed line shows the index for occupation (I_c^{Occ}).

and women—and that is what we see for both majors and occupations in Figure 4. For the 1950 birth cohort, women sorted into majors that reduced their potential wage by 12.5 percent relative to their male counterparts. Notably, like the time series pattern in the Duncan-Duncan segregation index, this potential wage index diverges slightly for the most recent birth cohorts. Even for the 1990 cohort, women are systematically sorted into majors associated with per-hour wages that are 9.5 percent lower than men. Figure 4 shows strong convergence in occupational segregation as measured by the potential wage index for occupations—the dashed line. College women from the 1950 birth cohort were in occupations that systematically had potential wages that were 14 percent lower than the occupations of their male counterparts. The potential wage gap in occupations fell to 10 percent for the 1990 cohort.

Collectively, these results highlight four facts about gender differences in undergraduate majors and occupations. First, the gap in potential wages based on major has declined somewhat across cohorts. Second, even for the most recent set of college graduates, a large gender gap in potential wages based on major still exists. Third, the gender gap in potential wages based on occupation is larger for all birth cohorts than the gender gap in potential wages based on major (the dashed line is always below the solid line). Finally, among college graduates, the convergence in the gap in potential wages based on major is of the same magnitude as the convergence in the gender gap in potential wages based on occupation.

Mapping Majors to Occupations

Our discussion to this point has documented sorting patterns in majors independent of sorting patterns in occupations. There are inevitable linkages between major and occupation. For example, college-educated nurses are more likely to be drawn from a pool of undergraduate Nursing or Science majors than Humanities majors. However, the overlap is not complete and those who complete the same college major need not end up in the same occupation. In this section, we explore the connections between pre-market and market specialization by documenting gender differences in the mapping between undergraduate major and occupation. We find that, conditional on major, women systematically sort into occupations with lower potential wages relative to men from the same cohort.

Occupational Mapping Patterns within Majors

Empirically, we find motivation for the idea of forking paths from majors to occupation in Figure 4 discussed above. The potential wage index based on occupation (dashed line) is consistently lower than the potential wage index based on major (solid line). This implies that conditional on being sorted into the same major as men, women systematically work in lower-pay occupations.

Table 1 shows the broad occupational distribution for men and women born between 1968 and 1977 for selected broad major categories. We summarize the occupational distribution by showing the four most common occupations associated with each major. Unlike undergraduate major, occupation may vary over an individual's lifetime. Thus, for Table 1 we measure occupations for the birth cohort from 1968 to 1977, at ages 38–47, using data from the 2014–2017 American Community Survey.

There are clearly large differences in occupational sorting between men and women who majored in the *same subject*. For example, among Education majors in the 1968–1977 birth cohort, 68 percent of all women and just 50 percent of all men who majored in Education worked as Teachers. Such patterns are robust across all birth cohorts. Again, for Education majors, 72 percent of women from the 1978–1987 birth cohort, 63 percent of women from the 1958–1967 birth cohort, and 52 percent of women from the 1948–1957 birth cohort worked as Teachers. The comparable numbers for men from the various birth cohorts were 58 percent, 43 percent, and 28 percent respectively.

Importantly, Table 1 shows evidence of rank effects. In all the broad majors in Table 1, women are less likely to be Executives and Managers and more likely to work

Table 1

Gender Differences in Occupation for Selected Majors, 1968–1977 Birth Cohort

Panel A. Education majors

	Teachers	*Executive/ Manager*	*Sales*	*Administrative Support*	$HHI_{g,c}^{Major}$
Men	0.50	0.18	0.06	0.03	0.29
Women	0.68	0.09	0.03	0.07	0.48

Panel B. Nursing/Pharmacy

	Nurses/ Health	*Executive/ Manager*	*Sales*	*Health Technicians*	$HHI_{g,c}^{Major}$
Men	0.46	0.15	0.07	0.06	0.25
Women	0.63	0.09	0.03	0.05	0.41

Panel C. Social Sciences

	Executive/ Manager	*Sales*	*Lawyers/ Judges*	*Administrative Support*	$HHI_{g,c}^{Major}$
Men	0.26	0.13	0.11	0.06	0.11
Women	0.20	0.07	0.08	0.13	0.10

Panel D. Business

	Executive/ Manager	*Sales*	*Accountant/ Underwriter*	*Administrative Support*	$HHI_{g,c}^{Major}$
Men	0.31	0.18	0.12	0.07	0.16
Women	0.24	0.11	0.17	0.18	0.14

Panel E. Engineering

	Executive/ Manager	*Engineer*	*Other Technicians*	*Architects/ Civil Engin.*	$HHI_{g,c}^{Major}$
Men	0.28	0.23	0.09	0.08	0.16
Women	0.27	0.18	0.05	0.07	0.13

Panel F. Biology/Life Sciences

	Physicians	*Executive/ Manager*	*Scientists/ Actuaries*	*Sales*	$HHI_{g,c}^{Major}$
Men	0.26	0.16	0.10	0.06	0.12
Women	0.18	0.14	0.08	0.04	0.10

Panel G. Physical Sciences

	Executive/ Manager	*Scientists/ Actuaries*	*Physicians*	*Sales*	$HHI_{g,c}^{Major}$
Men	0.20	0.15	0.08	0.07	0.09
Women	0.15	0.10	0.07	0.06	0.08

Panel H. Psychology

	Executive/ Manager	*Teachers*	*Sales*	*Psychologists/ Social Workers*	$HHI_{g,c}^{Major}$
Men	0.21	0.11	0.11	0.09	0.09
Women	0.16	0.21	0.06	0.13	0.11

Note: Table 1 shows the occupational distribution of men and women born between 1968 and 1977 for different majors. We use both broad major categories and broad occupational categories for this table. Each panel shows a different undergraduate major. The cells of the panel show the fraction of men (women) who majored in that occupation who subsequently worked in different broad occupations in the 2014–2017 American Community Survey. For each major, we show the four largest occupational categories based on where men with that major in that age range worked.

in Administrative Support occupations. Men from the 1968–1977 birth cohort who majored in Education are twice as likely as women (18 percent versus 9 percent) to be Executives or Managers—including principals and superintendents. Women who majored in Education are more than twice as likely as men (7 percent versus 3 percent) to work in Administrative Support roles—including teachers' aides, administrative assistants, and office supervisors. This effect is smallest among Engineering majors, where men and women are almost equally likely (28 percent versus 27 percent) to be Executives and Managers. Even in this case, male Engineering majors are more likely than female Engineering majors to work as Engineers (23 percent versus 18 percent).

Occupational Mapping Patterns across All Majors

Even when men and women have completed the same college major, they systematically end up in different occupations. Here, we explore whether women subsequently sort into a broader (less concentrated) or narrower (more concentrated) set of occupations than men with the same major. We find that women sort into a narrower set of occupations when 1) the major is female-dominated and 2) the major has a lower level of potential income. Throughout, we highlight that these patterns are consistent with men systematically moving towards higher earning occupations relative to women conditional on a given undergraduate major.

To summarize these patterns, we create a cross-occupation Herfindahl-Hirschman Index for each gender g and cohort c for every major. We calculate this index by: 1) calculating the share employed in an occupation of the relevant gender-major-cohort group; 2) squaring those shares; and 3) summing up over all occupations. For example, for a given cohort of History majors, let's say that 50 percent of the male History majors work in occupation A and 50 percent work in occupation B. Let's say that 95 percent of the female History majors work in occupation A and 5 percent work in occupation B. For men, the Herfindahl-Hirschman Index would equal 0.5; for women, it would equal 0.91. Our occupational concentration index ranges from 0 to 1, and higher levels imply that occupational sorting is more concentrated.[10]

Which majors differ the most or least by gender with respect to occupational concentration? Returning to Table 1—that is, the cohort born in 1968–1977 and then evaluated by their occupation from ages 38 to 47—the broad major categories with the largest gender difference in terms of occupational concentration are: 1) Education, Administration, and Teaching, 2) Nursing, Medical, and Health Services, 3) Public Affairs, Policy, and Social Work, 4) Construction Services, and 5) Criminal Justice and Fire Protection. For the Education, Administration, and Teaching; Nursing, Medical, and Health Services; and Public Affairs, Policy, and

[10] Formally, we define the index: $HHI_{g,c}^{Major} = \sum_o^O (s_{g,c}^{o|m})^2$ where $s_{g,c}^{o|m}$ is the share of group g from cohort c working in occupation o conditional on having major m. For each gender and cohort who matriculate with a given major, $\sum_o^O s_{g,c}^{o|m} = 1$ where O is the total number of potential occupations. This index measures the occupational concentration for individuals in each major separately by gender.

Social Work majors, women sort into a narrower set of occupations than their male peers. For the Construction Services and Criminal Justice and Fire Protection majors, men sort into a more concentrated set of occupations. The majors with the most similar occupational concentration between men and women are: 1) Environment and Natural Resources, 2) Business, 3) Social Sciences, and 4) Physical Sciences.

In Figure 5, we compute a Herfindahl-Hirschman index using the 134 detailed major categories focusing again on the 1968–1977 birth cohort.[11] Figure 5 highlights two notable patterns pertaining to gender differences in occupational concentration using the detailed major categories. Panel A plots the relationship between the extent to which the major is female-dominated (x-axis) and gender differences in occupational concentration for graduates with that major. In female-dominated majors, women systematically sort into a narrower set of occupations than men. Likewise, in male-dominated majors, men sort into a narrower set of occupations than women. Thus, panel A suggests strong relationship between the extent to which a major is dominated by one gender and the extent to which that gender sorts into a narrower set of occupations.

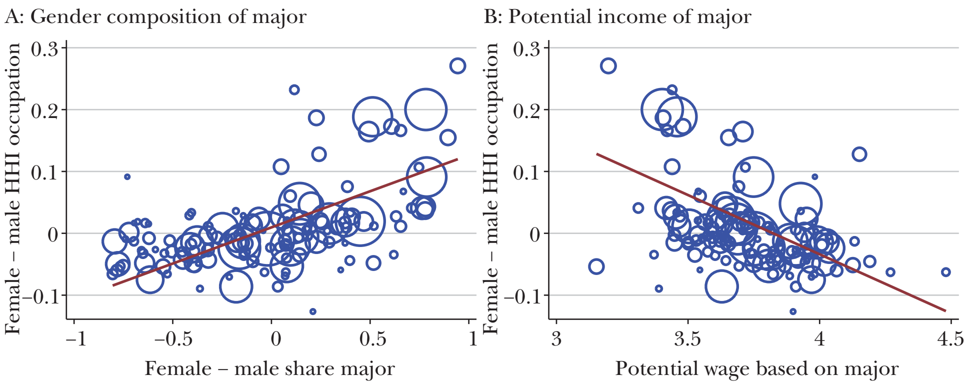
For example, consider the Elementary Education major that occupies the upper right quadrant of Figure 5. Women outnumber men in the major by 77 percentage points. The wedge in the Herfindahl-Hirschman index between women and men who major in Elementary Education is 0.15, which means that women with this major are much more likely to matriculate into a narrower set of occupations than comparable men. Consistent with the patterns for broad major categories in Table 1, almost all women who major in Elementary Education become teachers while more of the men become executives/managers (such as school principals). As an alternative example, consider the Theology major that occupies the lower left quadrant of Figure 5, Panel A. Male Theology majors outnumber women by 53 percentage points. The female-to-male difference in the Herfindahl-Hirschman Index is equal to –0.10 for Theology majors, which means that men who major in Theology are much more likely to matriculate into a narrower set of occupations than comparable women.

The anecdotal patterns underlying the results in Table 1 and panel A of Figure 5 suggest that women tend to sort into lower-paying occupations conditional on major. Panel B of Figure 5 provides additional evidence of this pattern. In particular, the figure highlights a strong and significant negative relationship between the potential income of a major, as we defined it above, and the female-to-male difference in occupational concentration for individuals graduating with that major. When the potential returns to a major are low, men disperse into a wider set of occupations than women, presumably to avoid the wage penalty of a low-return major. When the potential returns to a major are high, however, men sort into a *narrower* set of occupations than women. This is consistent with

[11] When computing the concentration measures for detailed majors, we still measure occupational sorting using our broad occupation categories.

Figure 5

Cross Major Variation in Within-Major Gender Differences in Occupational Concentration, 1968–1977 Birth Cohort



Source: Data shown only for the 1968–1977 birth cohort. Both panels include a fitted regression line. The slopes of the regression lines are 0.12 (standard error = 0.011) and –0.19 (standard error = 0.027), respectively.

Note: These figures show cross-major variation in the Herfindahl-Hirschman index as a function of how female-dominated is the major (panel A) and average major potential income (panel B). See text and the online Appendix for additional details. Each observation in both panels is a detailed major. We use broad occupation categories.

women migrating to lower income occupations (relative to men) when the potential income of a major is high.

For example, consider the Health and Medical Preparatory Programs that is the detailed major with the highest potential wage in the lower right quadrant of panel B. The female-to-male difference in occupational concentration is equal to –0.09, which means that men in this major are more likely to matriculate into a narrower set of occupations than comparable women. Many more men in this major end up being doctors relative to women in this major. The results in this section so far: 1) document that conditional on major men and women sort into different occupations; and 2) provide suggestive evidence that women systematically sort into lower earning occupations. In the next sub-section, we provide more specific evidence for the second result.

Cross-Cohort Patterns in Occupational Mapping

A shortcoming of our above occupational concentration index is that, like the Duncan-Duncan index, it does not take the earnings rank of occupations into account. Yet rank effects are quite evident in the descriptive exercise in Table 1. Consider two scenarios: in one, all male and female Finance majors sort into the same high-paying occupation: Executive or Manager. In the other scenario, all male Finance majors sort into the Executive or Manager occupation and all female

Finance Majors sort into the Administrative Support occupation. In both scenarios, men and women will have the same value of occupational concentration ($HHI_{g,c}^{Major}$). In other words, the concentration index is not in a price space and therefore does not account for men and women with the same major sorting into occupations with different potential wages.

In order to account for the true economic significance of differential major-to-occupation sorting patterns by gender, we introduce a summary statistic, $I_c^{Occ|m}$, that measures the gender gap in *potential occupational wages* for men and women graduating with the same major.[12] Consistent with this exercise throughout this paper, a potential wage refers to the log wages of native-born, White men, age 43 to 57 with strong attachment to the labor market. In the case of the potential occupational wage ($I_c^{Occ|m}$), we calculate the potential wage within an occupation conditional on matriculating with a given major. Each birth cohort, major, and gender in our sample has its own value of this index. Within a cohort and major, if these values differ by gender, it means that conditional on major, men and women have sorted into occupations with a different earnings ordering (rank). A negative value of the index implies that women with a certain major are systematically working in occupations with lower potential wages relative to men from the same major and cohort.

To build intuition, consider an example with a single major: Let's assume 70 percent of male Zoology majors work in occupation *A* and 30 percent work in occupation *B* while 30 percent of female Zoology majors work in occupation *A* and 70 percent work in occupation *B*. Recall, these distributions would yield the same concentration index for occupation. However, let's say the potential wage for occupation *A* (in logs) is 4.79 (the potential wage for the highest paid broad occupation in our sample–Physicians) and the potential wage for occupation *B* (in logs) is 2.71 (the potential wage for the lowest paid broad occupation in our sample—Housekeeping). In this example, the index would have a value of –0.83. This implies women Zoology majors choose occupations that are associated with 83 percent lower earnings on average relative to male Zoology majors.

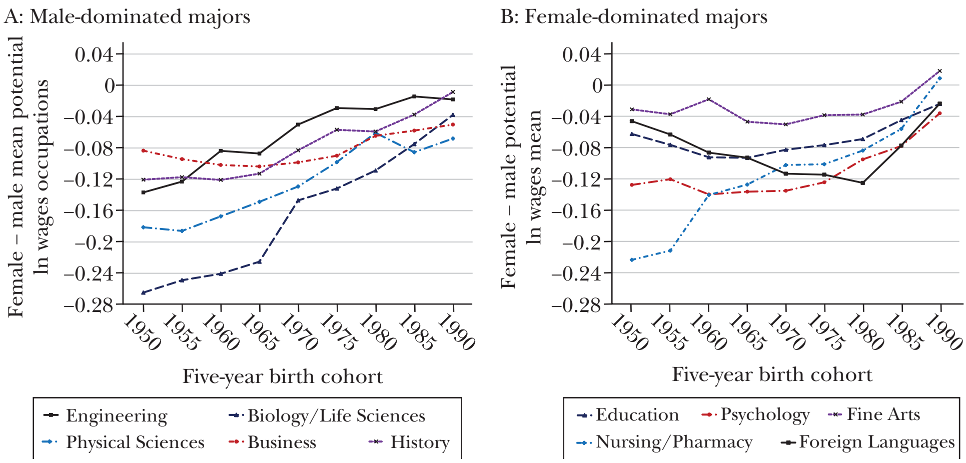
Conditional on graduating with the same major, do women systematically sort into occupations with lower potential wages? Figure 6 provides an initial answer to this question by examining the index value for the same broad categories of majors and cohorts shown earlier in Figure 1. The *y*-axis—the female-male difference in potential wages as determined by occupation conditional on major—contains mostly negative values, implying that for both male-dominated majors (panel A) and female-dominated majors (Panel B), women in each birth cohort are sorted

[12] As in the previous indices, we subtract the weighted sum for men of the potential wage as determined by occupation from the weighted sum for women of the potential wage as determined by occupation. The weights are the shares of men and women, respectively, who work in each occupation conditional on having majored in the same subject. Formally, we define our index as:

$$I_C^{Occ|m} = \sum_{m=1}^{M} \left(s_{female,c}^{Occ} \,|\, m\right) \bar{Y}_{male}^{Occ} - \sum_{m=1}^{M} \left(s_{male,c}^{Occ} \,|\, m\right) \bar{Y}_{male}^{Occ}$$

where $s_{g,c}^{Occ} \,|\, m$ is the share of gender *g* choosing occupation *Occ* conditional on major m from cohort *c*.

Figure 6
Within-Major Gender Differences in Potential Wage by Occupation, by Gender and Cohort



Source: 2014–2017 American Community Survey using the log wages of native-born, White men aged 43–57 with strong attachment to the labor market who work in a given occupation.
Note: These figures show the trends in the occupational sorting based on the occupation's potential log wage conditional on having graduated with a given major ($I_c^{Occ|m}$). Panel A are male-dominated majors. Panel B are female-dominated majors. As with Figure 4, occupational potential log wage is computed in the 2014–2017 American Community Survey using the log wages of native-born, white men 43–57 with strong attachment to the labor market who work in a given occupation.

into lower pay occupations conditional on major. There is evidence of convergence across cohorts. Take, for example, the Engineering major (solid line, panel A). For the 1950 birth cohort, the gender gap in potential wages based on occupation was large—female Engineering majors sorted into occupations that, on average, had potential wages that were 14 percent lower than their male peers. For the 1990 birth cohort, this difference fell to a 2 percent difference. Business majors saw this gap reduced by almost one-half although they experienced less convergence than Engineering majors.

The gender convergence in occupation within majors is seen in many but not all of the occupations in Figure 6. Many of the historically female-dominated majors like Education, Foreign Languages, and Fine Arts (panel B) saw more modest convergence across cohorts in the occupations taken by women relative to men (as measured by potential wages) conditional on major. Collectively, the patterns in Figure 6 highlight that for many broad majors, women sort into occupations with lower potential earnings relative to men, conditional on major, and that the extent of such differential sorting has diminished somewhat for recent cohorts.

The cross-cohort trends in Figure 6 are somewhat limited in that they describe just a few broad major categories in a taxonomic fashion by name and gender endowment (panel A versus panel B). In the online Appendix, we discuss in detail an exercise both to show these patterns for all majors in our sample and to provide a meaningful economic ordering of majors. In particular, we show the mapping to occupations, conditional on major, for two cohorts of college graduates who are likely to be settled into their careers: the 1955 birth cohort and the 1975 birth cohort. We know from earlier in the paper that compared to the 1955 cohort, women in the 1975 cohort have sorted into majors that were more similar to men. In this exercise, we show that compared to the 1955 birth cohort, women in the 1975 birth cohort also work in occupations that are more similar to their male peers *conditional on major.*

This convergence is non-trivial. For example, conditional on major, women from the 1955 birth cohort sorted into occupations that had potential earnings roughly 11 percent less than otherwise similar men. By the 1975 cohort, women sorted into occupations—conditional on major—that earned roughly 9 percent less than otherwise similar men. The convergence is driven by cross-cohort changes in occupation among those majoring in the highest paid majors. This change in the mapping of majors to occupations is one of the key findings of the paper.[13]

It is important here to note that our data from the 2014–2017 American Community Survey observes birth cohorts at different points in the life cycle. Recall that major in our study refers to undergraduate major of college graduates, not current college students. As such, major is fixed and will not change over time for an individual. In contrast, occupation is likely to change over one's life cycle. Because occupation is dynamic over the life cycle and increasingly dynamic across generations, this will complicate the interpretation of cross-cohort differences in occupation-based results. This limitation is particularly salient when considering the evolution of occupational sorting across cohorts. For this reason, the discussion in this section focused on birth cohorts who are likely settled in their occupations.

Analyzing the Wage Gap among College Workers

How much of the gender gap in college-graduate wages can be explained by controlling for both undergraduate major and for current occupational sorting? How have these relationships evolved over time? Previous scholarly work has grappled with these questions. In a classic reference, Brown and Corcoran (1997) use data from the 1984 Survey of Income and Program Participation (SIPP) and the National Longitudinal Study of High School Class of 1972 (NLS72) to document how coursework differences between men and women are associated with gender wages gaps. For

[13] In the online Appendix, we discuss a mapping result in an hours-worked space. In doing so, we show systematic sorting of women into occupations with lower potential hours worked than their male peers who graduated with the same major. In particular, over all majors and across all cohorts, conditional on major, women are in occupations that have a work requirement (based on male hours) that is about 3 percent less than comparable men. There is little trend in this gap across cohorts.

older cohorts born prior to 1960 (and thus, prior to the female overtaking in college completion), they find that undergraduate major accounts for 8 or 9 percentage points of the 20 percentage point college gender wage gap (where majors are divided into 20 broad categories). In related work, Loury (1997) uses data from the National Longitudinal Study of 1972 and the High School and Beyond Senior Cohort (Class of 1980) to document that controlling for grade point average and four broad major categories reduces the gender wage gap. Black et al. (2008) use data from the 1993 National Survey of College Graduates to examine the extent to which pre-labor market factors—including broad undergraduate major—explain differences in wages across various race-gender groups of college graduates. Within a broader analysis of factors contributing to the gender wage gap such as psychological attributes and demands for flexibility, Bertrand (2017) uses data from the 2012–2015 American Community Survey to document cross-cohort convergence in potential wages based on degree attainment and major.

Returning to our sample from the 2014—2017 American Community Survey, we explore these issues, both for all cohorts pooled together and also separately by 10-year birth cohorts. The latter analysis lends itself to a decomposition exercise to assess how much of the change in gender wage gaps can be explained by changes in the distribution of undergraduate majors and occupations. Specifically, we estimate regressions of the following form:

$$\ln(Wage)_i = \alpha + \beta Female_i + \delta_m Major_i + \delta_o Occ_i + \Gamma X_i + \epsilon_i$$

where $\ln(Wage)_i$ is a measure of individual i's log wage and $Female_i$ is a dummy variable equal to 1 if the individual is female. Our estimated variable of interest is β that measures the gender gap in log wages. The variables $Major_i$ and Occ_i are summary measures of the individual's chosen undergraduate major and observed occupation. We summarize an individual's major and occupation with the potential log wage variables $\bar{Y}_i^m$ and $\bar{Y}_i^o$. In all specifications, we include a vector of demographic controls summarized in the vector X_i. Specifically, we control for five-year birth cohort, race, state of residence, educational attainment beyond a bachelor's degree, survey year, and marital status. Standard errors are clustered by state of residence.[14]

Table 2 summarizes the basic results. In the top panel we show results pooling together individuals from all the birth cohorts in our sample. In column 1, we use only demographic controls for highest degree completed, age, race, and state of residence to estimate the gender gap in wages for all college-educated cohorts. The gap is estimated at 23 percent, meaning college women in our sample earn about 77 cents on a dollar compared to college men.

[14] In the online Appendix, we report results from an alternate specification where we do not include demographic controls. We also report results from two alternate specifications where we aggregate majors and occupations to broader categories and instead include dummies for each broad major and occupation category. These exercises yield results that are very similar to those in Table 2.

Table 2
Majors, Occupations, and Gender Gaps in Wages

Panel A. log wage regressions, pooled cohorts

	log wages			
Variable	*(1)*	*(2)*	*(3)*	*(4)*
$Female_i$	−0.233	−0.158	−0.143	−0.114
	(0.006)	(0.004)	(0.004)	(0.003)
$\bar{Y}_i^m$		0.807		0.408
		(0.015)		(0.012)
$\bar{Y}_i^o$			0.757	0.677
			(0.011)	(0.009)
Controls	Yes	Yes	Yes	Yes
R^2	0.22	0.27	0.36	0.37

Panel B. log wage regressions, separately by cohort

	1958–1967 birth cohorts			*1978–1987 birth cohorts*		
Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
$Female_i$	−0.322	−0.198	−0.168	−0.155	−0.093	−0.065
	(0.008)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
$\bar{Y}_i^m$			0.411			0.443
			(0.016)			(0.010)
$\bar{Y}_i^o$		0.909	0.823		0.599	0.513
		(0.015)	(0.012)		(0.008)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.13	0.32	0.33	0.13	0.25	0.27

Note: Table 2 shows estimates from regression (3). See text and online Appendix for additional details. Sample size for panel A columns 1–4 is 1,135,196. Sample size for panel B columns 1–3 is 266,674. Sample size for panel B columns 4–6 is 307,053.

In the rest of the analysis, we consider the effects of specialization on the gender wage gap. The $Female_i$ coefficient in column 3 reports the gender wage gap when only controlling for demographic characteristics and occupation. This gap is estimated at 14.3 percent or about 86 cents on a dollar. Consistent with the existing literature on the gender wage gap, this estimate tells us that market specialization (occupation) matters.

How much predictive power does major have above and beyond controlling for just occupation? In column 4, we report results from a regression that fully controls for demographics, occupation, and college major. Adding undergraduate major further reduces the gender wage gap to 11 percent—a reduction of about one-half from our model with no specialization controls in column 1. Adding major to our model accounts for an additional 3 percent of the college gender wage gap above and beyond a model that only accounts for market specialization. In this calculation, major independently accounts for one-quarter of the reduction in the college gender wage gap. In other words, college women earn 89 cents on a dollar earned by college men who have sorted into similar majors and occupations.

How has the gender gap evolved across generations of US college graduates? In the bottom panel of Table 2, we compare recent (1978–1987) birth cohorts to older cohorts (1958–1967). In the most basic models that control only for demographics, we see a reduction in the gender wage gap among college graduates from 32.2 percent for the older cohort to 15.5 percent for the younger birth cohort. These results can be found in columns 1 and 4. We find that in our fully controlled models, there is a further reduction in the gender wage gap between the older cohort (column 3) and the younger cohort (column 6). Controlling for demographics, occupation, and major, the 1958–1967 birth cohort had a 16.8 percent gap, or about 83 cents on a dollar. By the 1978–1987 birth cohort, this gap fell to 6.5 percent or about 94 cents on a dollar. The role of college major in these cohorts increased modestly, but there was a sharp reduction in the importance of occupation. Overall, we find that controlling jointly for major and occupation explains roughly 40 percent of the cross-birth cohort decline in the wage gap.

In the 1950 birth cohort, men completed bachelor's degrees at a higher rate than women, but for all subsequent birth cohorts, women surpassed men in bachelor's completion. In a counterfactual exercise, we equalize the male and female distributions. Upon doing this, we can bound the above estimates of the college gender wage gap. If the number of male and female college graduates in each five-year cohort was the same, the gender wage gap after controlling for simple demographics was between 21 and 27 log points. Once we control for demographics, occupation, major and equalize the distributions, the college gender wage gap was between 11 and 12 log points.

In additional work, we conducted a wage decomposition exercise to better understand the power of our explanatory variables within cohort. A detailed discussion including a table of results can be found in the online Appendix. We find occupational specialization explains the largest share of the gender wage gap for college graduates. For example, occupation explains 36.9 percent of the gender wage gap in the youngest cohort (1978–1987). Sorting by major is also important and explains 27.9 percent of the gender wage gap in that same cohort. Notably, human capital attainment above and beyond a bachelor's degree (such as a graduate degree) explains considerably less of the college gender wage gap. These results suggest that properly accounting for human capital decisions above and beyond schooling attainment and occupational specialization is centrally important in understanding the causes of the gender wage gap among the highly skilled.

Separately, we document that undergraduate major does not have any effect on extensive margin labor market participation for college graduates. While undergraduate major is informative about gender wage differentials, it is not informative with respect to explaining extensive margin gender differences in labor supply. However, we document heterogeneous effects by gender on intensive margin participation.

Specifically, we find that conditional on undergraduate major, women sort into occupations with lower potential annual hours worked than men.[15]

Discussion

A gendered specialization of human capital and labor has primitive roots. A division of labor in the early home-based economy was largely influenced by biological differences between the sexes, particularly with respect to manual tasks. As workers wandered beyond the home and field to factory employment, many tasks remained manual, and biological differences continued to dictate a division of labor. We see historical evidence of this in US manufacturing at the end of the 19th century: women worked in mostly precision manufacturing occupations such as tobacco, textile, apparel, paper, and rubber, with men dominating most other manufacturing occupations including heavy machinery manufacturing used in the production and fabrication of metal.[16] Modern US history has seen improved knowledge of and access to family planning, a subsequent decline in fertility, and a sustained growth in occupations that require relatively fewer manual tasks and relatively more cognitive tasks (for discussion, see Bailey 2006; Blau, Ferber, and Winkler 2014; Autor and Dorn 2013). Such factors would erode a physicality-based male comparative advantage in the labor market.

Further, as technology was altering the occupational landscape with respect to task demands, changes were afoot with respect to college education. For the US birth cohorts from 1870 to 1910, the male and female college graduation rates were close and paralleled in trend. For the birth cohorts from the mid-1910s to the mid-1950s, the male series pulled away from the female series creating a sustained period of male comparative advantage on average with respect to tasks requiring a college degree. The late 1950s birth cohorts saw a male–female convergence in college completion with women eventually surpassing men, giving cohorts of younger women both in the United States and globally a comparative advantage on average in a global labor market that was experiencing upskilling—a trend that has not reversed (Becker, Hubbard, and Murphy 2010; Charles and Luoh 2003; DiPrete and Buchmann 2006; Goldin, Katz, and Kuziemko 2006; Jacob 2002).

The shifting comparative advantage also seems to have affected women's investments while at university. Using the 1993 National Survey of College Graduates, Black et al. (2008) report that among women with a bachelor's degree born in the 1930s, 38.3 percent majored in Education and 12 percent majored in the Humanities. For women born between 1960 and 1965, only 14.8 percent majored in Education and only 6.8 percent majored in the Humanities. As we documented

[15] A full discussion of these results can be found in the online Appendix.

[16] According to the 1890 US Manufacturing Census, tobacco, textile, apparel, paper, and rubber were 31.9 percent, 40.8 percent, 60.2 percent, 39.9 percent, and 48.8 percent female respectively (Goldin 1990, p. 80).

in Figure 4, women's majors converged in terms of potential wages until about the 1975 birth cohort and then experience a puzzling reversal. Since then, the gap in potential wages between women and men has increased, although it is important to keep in mind that in this era relatively more women graduated from college. During this time, overall college enrollments grew, with much of that growth coming from increased female enrollments. Thus, as college enrollments ballooned, we may expect marginal students to select less challenging majors, which may be associated with lower potential wages.

For college graduates, success in the labor market is not determined by pre-market specialization (college major) alone, but is also influenced by specialization that happens in the market (occupation). Unlike field of study, occupational specialization requires workers and firms to match on the assignment. Using a potential wage index for occupation constructed in the same way as the one for major, we document strong male–female convergence in potential wages. Importantly, we see a flattening but do not see a reversal in the trend as we do in the major series.[17]

Despite the convergence in potential wages based on both occupation and major, the gender wage gap among college graduates remains substantial. Controlling for major and occupation sorting explains roughly 60 percent of this gap. These patterns are a highly salient topic for future research. The sustained importance of major in explaining the gender wage gap—a specialization outcome set in motion before workers even enter the market and face market frictions such as tenure disruptions due to family demands or employer discrimination—highlights the need to better understand the mechanisms driving pre-market investments. Whether women choose a major in anticipation of future family demands, based on individual preferences, under the burden of restrictive social norms, or for any other reason may be best explored in an experimental setting or with access to data where preferences and not solely outcomes are observed.

In closing, our thoughts return to the home sector. As work outside the home has evolved to involve fewer manual tasks, there should be a subsequent narrowing of female comparative advantage in home production. If this comparative advantage does not completely disappear, we may expect sustained gender differences with respect to specialization before and in the market, which will result in disparities in labor market outcomes. In a model in which men and women are equally productive in the market but women have a comparative advantage at home, Lazear and Rosen (1990) predict tougher promotion standards for women than men. If there is any investment feedback effect, we may anticipate different educational and specialization decisions by workers who have a smaller probability of promotion—in this case, female workers. Recent empirical evidence, summarized in Cortés and Pan

[17] For those who wish to explore the literature documenting both changes in occupational sorting by gender over time and the contribution of occupation to gender labor market disparities, useful starting points include Blau, Ferber, and Winkler (2014); Cortés and Pan (2018); Hsieh et al. (2019), and the references cited therein.

(2020), points to sustained gender wage gaps among all workers. With respect to the influence of home production, particularly child-rearing, the paper reports a long-term earnings loss of 39 percent for working mothers relative to the birth year of their first child. Understanding the relationship between anticipated demands at home and specialization decisions that occur both before and in the market is central to understanding gender disparities in outcomes.

■ *We thank seminar participants at the University of Chicago and conference participants and discussants at SOLE/EALE and the NBER Summer Institute Gender Studies program for helpful comments. A longer version of this paper circulated as "A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap."*

References

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–97.

Bailey, Martha J. 2006. "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply." *Quarterly Journal of Economics* 121 (1): 289–320.

Becker, Gary S., William H.J. Hubbard, and Kevin M. Murphy. 2010. "Explaining the Worldwide Boom in Higher Education of Women." *Journal of Human Capital* 4 (3): 203–41.

Bertrand, Marianne. 2017. "The Glass Ceiling." Becker Friedman Institute for Research in Economics Working Paper 2018-38.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. "Gender Wage Disparities among the Highly Educated." *Journal of Human Resources* 43 (3): 630–59.

Blau, Francine D., Marianne A. Ferber, and Anne E. Winkler. 2014. *The Economics of Women, Men and Work*. London: Pearson.

Brown, Charles, and Mary Corcoran. 1997. "Sex-Based Differences in School Content and the Male–Female Wage Gap." *Journal of Labor Economics* 15 (3): 431–65.

Charles, Kerwin Kofi, and Ming-Ching Luoh. 2003. "Gender Differences in Completed Schooling." *Review of Economics and Statistics* 85 (3): 559–77.

Cortés, Patricia, and Jessica Pan. 2018. "Occupation and Gender." In *Oxford Handbook of Women and the Economy*, edited by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, 425–452. Oxford: Oxford University Press.

Cortés, Patricia, and Jessica Pan. 2020. "Children and the Remaining Gender Gaps in the Labor Market." NBER Working Paper 27980.

Dickson, Lisa. 2010. "Race and Gender Differences in College Major Choice." *Annals of the American Academy of Political and Social Science* 627 (1): 108–24.

DiPrete, Thomas A., and Claudia Buchmann. 2006. "Gender-Specific Trends in the Value of Education and the Emerging Gender Gap in College Completion." *Demography* 43 (1): 1–24.

Dorn, David. 2009. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." PhD diss. Verlag nicht ermittelbar.

Dorn, David. 2013. "ACS occ to occ1990dd." https://www.ddorn.net/data/occ2005_occ1990dd.zip, accessed (accessed July 24, 2019).

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes."

American Sociological Review 20 (2): 210–17.

England, Paula, and Su Li. 2006. "Desegregation Stalled: The Changing Gender Composition of College Majors, 1971–2002." *Gender and Society* 20 (5): 657–77.

Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* Oxford: Oxford University Press.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20 (4): 133–56.

Hsieh, Chang-Tai, Erik Hurst, Chad I. Jones, and Pete J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica* 87 (5): 1439–74.

Jacob, Brian A. 2002. "Where the Boys Aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21 (6): 589–98.

Lazear, Edward P., and Sherwin Rosen. 1990. "Male–Female Wage Differentials in Job Ladders." *Journal of Labor Economics* 8 (1): S106–23.

Loury, Linda Datcher. 1997. "The Gender Earnings Gap among College-Educated Workers." *Industrial and Labor Relations Review* 50 (4): 580–93.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "IPUMS USA: Version 9.0 [dataset]." Minneapolis, MN: IPUMS, 2019, 10, D10 (accessed July 24, 2019).

Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources* 48 (3): 545–95.

Journal of Economic Perspectives—Volume 35, Number 4—Fall 2021—Pages 249–255

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.

Smorgasbord

The Bank of International Settlements, in its *Annual Economic Report 2021* report, devotes a chapter to "CBDCs: an opportunity for the monetary system" (June 2021, https://www.bis.org/publ/arpdf/ar2021e.htm). "Central bank digital currencies (CBDCs) offer in digital form the unique advantages of central bank money: settlement finality, liquidity and integrity. They are an advanced representation of money for the digital economy. . . . The ultimate benefits of adopting a new payment technology will depend on the competitive structure of the underlying payment system and data governance arrangements. The same technology that can encourage a virtuous circle of greater access, lower costs and better services might equally induce a vicious circle of data silos, market power and anti-competitive practices. CBDCs

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.4.249.

and open platforms are the most conducive to a virtuous circle. . . . CBDCs built on digital identification could improve cross-border payments . . . .”

The McKinsey Global Institute has published “The economic state of Black America: What is and what could be” (June 2021, https://www.mckinsey.com/featured-insights/diversity-and-inclusion/the-economic-state-of-black-america-what-is-and-what-could-be). “Dismantling the barriers that have kept Black Americans from fully participating in the US economy could unleash a tremendous wave of growth, dynamism, and productivity . . . Today the median annual wage for Black workers is approximately 30 percent, or $10,000, lower than that of white workers . . . We estimate a $220 billion annual disparity between Black wages today and what they would be in a scenario of full parity, with Black representation matching the Black share of the population across occupations and the elimination of racial pay gaps within occupational categories. . . . The racial wage disparity is the product of both representational imbalances and pay gaps within occupational categories—and it is a surprisingly concentrated phenomenon. . . . The median Black household has only about one-eighth of the wealth held by the median white household. The actual dollar amounts are striking: while the median white household has amassed $188,000, the median Black family has about $24,000. . . . We estimate a $330 billion disparity between Black and white families in the annual flow of new wealth, some 60 percent of which comes from inheritances. . . . The gap in inheritances between Black and white recipients is some $200 billion annually . . . .”

The Credit Suisse Research Institute discusses “The global food system: identifying sustainable solutions” (June 2021, downloadable at https://www.credit-suisse.com/about-us-news/en/articles/news-and-expertise/sustainable-food-as-an-investment-opportunity-202106.html). “A sustainable global food system benefits human health as well as the global ecosystem. However, this is far from the reality at present as almost 700 million people are undernourished, while at the same time around 1.8 billion people globally are overweight or obese. The need to change appears obvious to us as the impact of malnutrition alone costs the global economy USD 13.6 trillion annually. . . . Malnutrition is not the only reason why the global food system needs to change. Food production and consumption already contribute well over 20% to global greenhouse gas emissions and account for more than 90% of the world’s freshwater consumption. . . . The likely growth in the world’s population to around ten billion people by 2050 coupled with a further shift in diets, especially across the growing emerging middle class, could increase emissions by a further 46%, while demand for agricultural land could increase by 49%.”

Edited E-books

Ruud de Mooij, Alexander Klemm, and Victoria Perry have edited a collection of 16 interrelated essays concerning *Corporate Income Taxes under Pressure: Why Reform Is Needed and How It Could Be Designed* (IMF, 2021, https://www.elibrary.imf.

org/view/books/071/28329-9781513511771-en/28329-9781513511771-en-book.xml). Narine Nersesyan lays out some of the overall issues in "The Current International Tax Architecture: A Short Primer." "When a business activity crosses national borders, the question arises as to where the profits resulting from that activity should be taxed. In principle, there are at least three possibilities for assigning a taxing right: • *Source:* the countries where production takes place • *Residence:* the countries where a company is deemed to reside • *Destination:* the countries where sales take place. The generally applied tax architecture for determining where profits are taxed is now nearly 100 years old—designed for a world in which most trade was in physical goods, trade made a less significant contribution to world GDP, and global value chains were not particularly complex. . . . The current international tax framework is based on the so-called '1920's compromise' . . . [in which] the primary right to tax active business income is assigned where the activity takes place—in the 'source' country—while the right to tax passive income, such as dividends, royalties and interest, is given up to the 'residence' country—where the entity or person that receives and ultimately owns the profit resides. The system has, however, evolved in ways that considerably deviate from this historic 'compromise,' and international tax arrangements currently rest on a fragile and contentious balance of taxing rights between residence and source countries . . . "

Ana Margarida Fernandes, Nadia Rocha, and Michele Ruta have edited an e-book of 16 essays titled *The Economics of Deep Trade Agreements* (CEPR Press, June 2021, https://voxeu.org/content/economics-deep-trade-agreements-new-ebook). From their Introduction: "Pascal Lamy, former Director General of the World Trade Organization (WTO), recently wrote: 'More than tariffs, trade agreements today are about regulatory measures and other so called 'non-tariff measures', that were once the exclusive domain of domestic policy-making. For these reasons, 'deep' trade agreements, as trade experts refer to this new class of agreements, are fundamentally different than the previous generation of trade agreements.' . . . Starting in the early 1990s, the number of PTAs [preferential trade agreements] has increased from 50 to more than 300 within three decades. While WTO rules still form the basis of most trade agreements, PTAs have in some sense run away with the trade agenda. . . . While the average PTA in the 1950s covered 8 policy areas, in recent years they have averaged 17 . . . At the same time, the number of commitments that governments have taken in trade agreements has largely increased, along with provisions requiring stronger transparency . . . "

Alain Samson has edited the *The Behavioral Economics Guide 2021* (Behavioral Science Solutions, 2021, https://www.behavioraleconomics.com/be-guide/the-behavioral-economics-guide-2021). It includes 15 short chapters by authors summarizing their recent work, along with a 30+ page glossary with short essays (and citations) on behavioral science concepts, starting with "action bias" and ending with the "zero price effect." John A. List contributes an introductory essay on "The Voltage Effect in Behavioral Economics:" "Indeed, most of us think that scalable ideas have some 'silver bullet' feature, i.e., some quality that bestows a 'can't miss' appeal. That kind of thinking is fundamentally wrong. There is no single quality that distinguishes ideas that have the potential to succeed at scale with those that do not do so. In this manner, moving from an initial research study to one that will have an attractive benefit cost profile at scale is much more complex than most imagine. And, in most cases, scaling produces a voltage drop—the original BE [behavioral economics] insights lose considerable voltage when scaled. The problem, ex ante, is determining whether (and why) that voltage drop will occur. . . . What this lesson inherently means is that scaling, in the end, is a weakest link problem: the endeavor is only as strong as the weakest link in the chain."

Interviews

David A. Price has interviewed "Ayşegül Şahin: On wage growth, labor's share of income, and the gender unemployment gap" (*Econ Focus*: Federal Reserve Bank of Richmond, Second/Third Quarter 2021, pp. 18–22, https://www.richmondfed.org/publications/research/econ_focus/2021/q2-3/interview). "What was striking about the Great Recession was its persistence. Everybody kept saying at the time that inflation is around the corner, the labor market is getting tighter, but it took a very long time for the labor market to heal. We are not seeing that this time. This was a very different shock. It was sharp, but it was transitory compared to the Great Recession. So the effect was great, but the recovery has been faster as well. I think that's the main difference. Another big difference is that the Great Recession was a big shock to the construction sector, and we are seeing the opposite now. We've been spending more time at our houses and people want to improve their houses and they want bigger houses. . . . But the biggest difference is the persistence. After the Great Recession, it took quits rates five or six years to recover. Today, the quits rate is already back to where it started from before the pandemic hit. . . . During the Great Recession, this aversion to quitting lasted for a long time. As a result, people were stuck in jobs that they were not necessarily happy about or they were not very productive at. But in this recession, quits rates bounced back quickly. One reason is because there are a lot of job openings; the second is that people want to go back and find jobs that they are better matched at."

Luis Garicano serves as interlocutor in *Capitalism after COVID: Conversations with 21 Economists* (June 2021, CEPR Press, https://voxeu.org/content/capitalism-after-covid-conversations-21-economists). As one example, here's Jesús Fernández-Villaverde in an interview called "Economists and the pandemic:" "One thing that I find disappointing for societies is that either you work eight hours a day or you work zero. Some societies, especially in the north of Europe, have made progress in terms of solutions, but it's something we should push very hard. . . . So, imagine we are in a society where we have more flexible forms of work. Thanks to Zoom, I can go from being expected to work eight hours a day, to being expected to work six hours and half hours, to say a random number. Then, it's much easier to reconcile work with family. . . . A lot of what we do is about coordination. I want to be in the office at 9:00 because someone else is going to be at there at that time.

You just need to have a focal point, and governments can help to coordinate us in good focal points. I can imagine many people who are aged 65 or 66 and are reluctant to work eight hours a day, but at the same time they are not very happy working zero hours a day. If we could get to a society where you can flexibly work four or five hours a day, maybe we could extend the working life of many people, contributing to GDP and helping us a lot to transition. . . . Thanks to telecommuting and Zoom, we may be able to do that much better than in the past."

The Centre for Development and Enterprise has published "Lant Pritchett in Conversation with Ann Bernstein" (June 2021, https://www.cde.org.za/wp-content/uploads/2021/06/Lant-Pritchett-in-conversation-with-Ann-Bernstein.pdf). "Development is a process that happens at the level of countries. The four transformations a country should make are: (1) to a productive economy, (2) to a capable state, so that it is able to do what it sets out to do, (3) to a government responsive to the needs and wishes of citizens, and (4) to a society where equal treatment of all before the law and of each other is a bedrock principle. I think those four characterise the transformation that takes a country from chaos and poverty to the levels of prosperity and well-being that we see in developed countries. . . . You often hear the phrase 'this or that isn't a panacea.' My argument is: national development is a panacea. If your country manages to undergo the four transformations of national development, then all problems get solved because that is a machinery for nominating and solving problems. Yet the current focus in development is on what I call 'kinky development,' which involves tinkering on the margins to help the poorest of the poor. That is the wrong focus. If you achieve national development, you will solve poverty and provide prosperity for the general population, whereas focusing on poverty alone often is at odds with getting you to desirable levels of prosperity. . . . No country has high levels of human wellbeing without having achieved national development; and every country that has high national development achieves very high levels of human well-being. So, the only path to high human well-being is through national development."

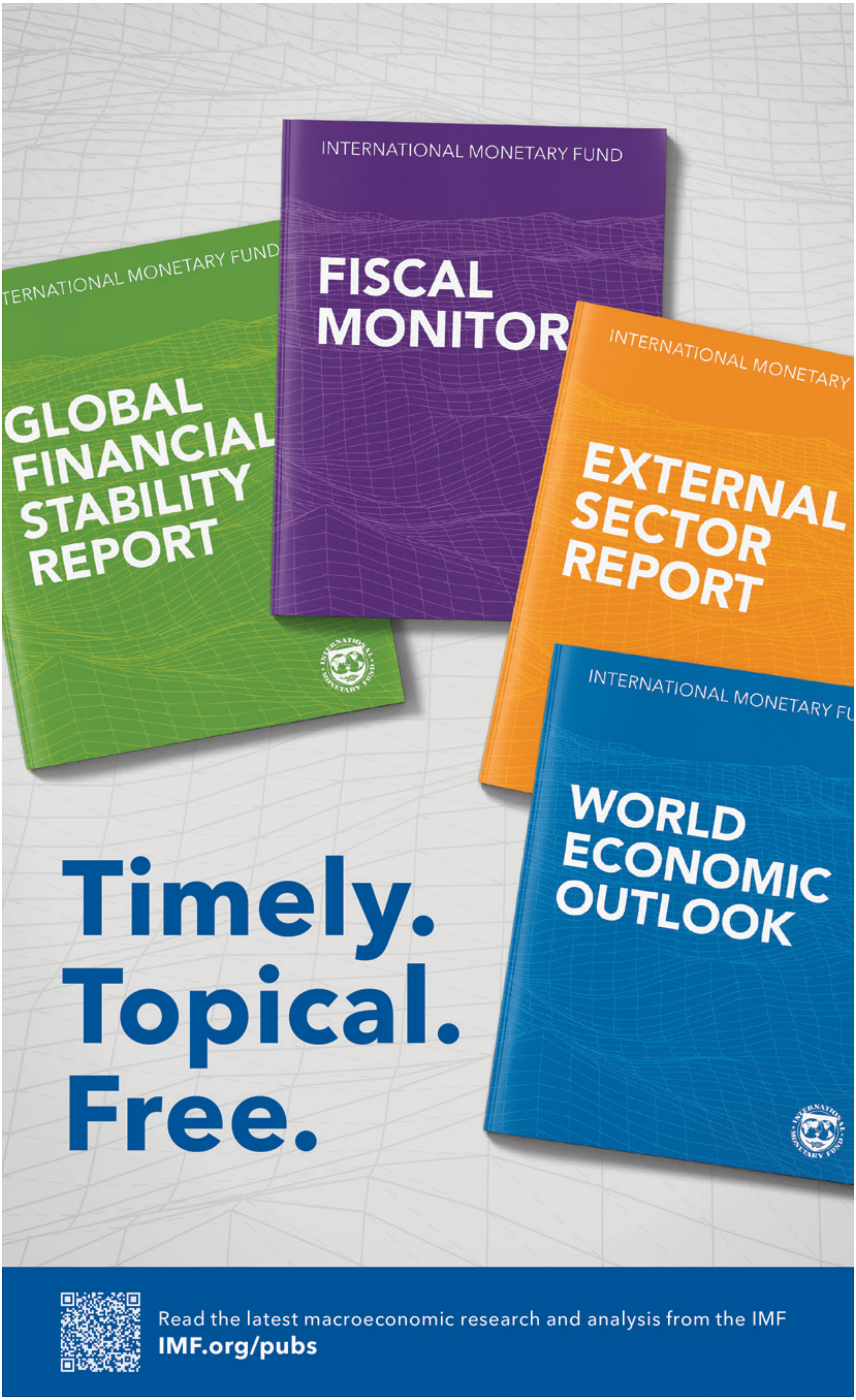
Michael Chui and Anna Bernasek of the McKinsey Global Institute discuss "Forward Thinking on technology and political economy with Daron Acemoglu" (July 14, 2021, https://www.mckinsey.com/featured-insights/future-of-work/forward-thinking-on-technology-and-political-economy-with-daron-acemoglu). "[I]f you look at the way that economists think about technology, it's this latent variable that makes you just more productive. But very few technologies actually do that. Electricity didn't make workers more productive. It made some functions in factories more feasible, and some few items more productive. . . . The example of spinning and weaving machinery that I gave, or the factory system, or, more recently, databases, software, robots, numerically controlled machines, they are mostly about replacing workers in certain tasks that they used to perform. . . . [I]n fact, one of the striking but very robust features of the last 40 years of economic development in the United States and the United Kingdom has been that many groups, especially low-education or middle-education men, have actually seen their earnings fall, some groups by as much as 25 percent, in real terms, since 1980. . . . In the traditional economics approach . . . it is something that doesn't really fit into this technology as augmenting framework. But when technology, at least in part, is about automation, replacing, displacing workers from their tasks, then this happens quite often. You can have productivity improvements—capital benefits, firms benefit, but workers, especially some types of workers, all workers overall can lose out in real terms. . . . [O]nce you go to this micro level, then the direction of technology, the future of technology looked at through the perspective of what type of technologies we're going to build on, that becomes much richer and much more interesting. . . . [W]e have to come back to a world in which we put as much effort in increasing human productivity, both in the tasks that they already produce, but also creating new tasks in entertainment, in healthcare. . . . There are many, many things ranging from judgment, social skills, flexibility, creativity, that humans are so much better at than machines. But we're not empowering them right now."

Bill Kerr interviews Erica Groshen in "Infrastructure: Upgrading the US labor statistics system" (Harvard Business School podcast, June 30, 2021, https://www.hbs.edu/managing-the-future-of-work/podcast/Pages/podcast-details.aspx?episode=19653416, transcript and audio). "[S]tate unemployment insurance agencies that, as part of running the program, collect worker wage records every quarter from every employer that lists the wages of workers for every month during that quarter. They also collect claims records from people who apply for claims. And these data are generally not available to BLS [Bureau of Labor Statistics] to augment or replace its current data collections. And that's basically a shame, because it would be quite useful for statistical purposes. And employers, of course, have to report the same or slightly different data to a number of different government agencies. Our economics statistics are also not as good as they could be as a consequence of this. UI [unemployment insurance] wage records include who the person's employer is and their earnings—that's what's in there. They should have job title, because that is closely associated with the person's occupation. . . . And this would enable us to track workforce supply and demand much more closely, make better projections about the future of work. You also would want the number of hours worked for the wages that are being reported so that you know if someone is full time or part time, so you can get hourly rates, and really follow that dimension on which wages vary. Another thing you want is the actual work location of the people. . . . And then, the last thing, particularly in these times of understanding demographic inequities—racial inequities, in particular, but also gender inequities, things like that—you want to have demographics so that you can track social justice issues and advances and understand how the world of work is affecting demographic outcomes. These data should also, of course, be curated—by which I mean, they have to clean them up so that you can really analyze them and made accessible to the statistical agencies, for particular with the BLS, so that they can create better statistics. You could get better, cheaper, and more-frequent program-policy evaluations so that policy makers could make better decisions."

Discussion Starters

Ellis P. Monk Jr., Michael H. Esposito, and Hedwig Lee discuss "Beholding Inequality: Race, Gender, and Returns to Physical Attractiveness in the United States" (*American Journal of Sociology*, July 2021, pp. 194–241, https://www.journals.uchicago.edu/doi/abs/10.1086/715141). "While a one-standard-deviation increase in ability is associated with 3%–6% higher wages, attractive or very attractive individuals earn 5%–10% more than average-looking individuals. Another study even finds that returns to perceived attractiveness unfold over the life course and are robust to a wide array of potentially relevant controls, such as educational attainment, parental background, personality traits, IQ, and so on. . . . [P]erceived physical attractiveness is a major factor of inequality and stratification regardless of one's race or gender. In fact, our analyses suggest that the magnitude of the earnings gap among White men along the perceived attractiveness continuum rivals that of the canonical Black-White wage gap and the attractiveness earnings gap among White women actually exceeds, in real dollars, the Black-White wage gap. . . . We find that while the returns to perceived physical attractiveness are similar for most race-by-gender combinations, the slope of the returns to perceived physical attractiveness is steepest among Black women and Black men. . . . Among Black women and Black men, the wage penalties associated with perceived physical attractiveness are also so substantial that, taken together, the earnings disparity between the least and most physically attractive exceeds in magnitude both the Black-White wage gap and the gender gap."

Mary Brooks and Paul Rosenzweig describe a modest resurgence in prediction markets in "Let's Bet on the Next Big Policy Crisis—No, Really" (Lawfare blog, July 13, 2021, https://www.lawfareblog.com/lets-bet-next-big-policy-crisis-no-really). "Metaculus offers a platform for a quasi-prediction market, in which the currency of exchange is prestige points, and anyone can submit a question for inclusion in the market. . . . [T]here is significant demand for internal corporate prediction markets and crowd-forecasting. Google, Ford, Yahoo, Hewlett-Packard, Eli Lilly and a number of other prominent corporations have operated or continue to operate a corporate market. . . . From 2011 to 2015, the Intelligence Advanced Research Projects Activity (IARPA)—the intelligence-minded sister of DARPA—ran the Aggregative Contingent Estimation (ACE). ACE was a project designed to 'dramatically enhance the accuracy, precision, and timeliness of intelligence forecasts . . . [by means of] techniques that elicit, weight, and combine the judgments of many intelligence analysts.' Today, IARPA still runs the Hybrid Forecasting Competition, which "develop[s] and test[s] hybrid geopolitical forecasting systems." . . . Kalshi—a San Francisco-based startup currently operating in beta—is the first fully regulated (CFTC-approved) prediction market. Because Kalshi is regulated, more significant amounts of money can be wagered than in many other markets, enabling them to build out a new asset class of events futures. The implications for this are obvious: An asset class like this could serve as an alternative or a supplement to more traditional insurance, allowing companies and individuals to hedge against crop failures, cyberattacks or floods."


INTERNATIONAL MONETARY FUND
FISCAL MONITOR
GLOBAL FINANCIAL STABILITY REPORT
EXTERNAL SECTOR REPORT
WORLD ECONOMIC OUTLOOK
Timely.
Topical.
Free.
Read the latest macroeconomic research and analysis from the IMF
IMF.org/pubs

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $24.00, $34.00, or $44.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

The following Statement of Ownership, Management and Circulation is provided in accordance with the requirements, as contained in 39 U.S.C. 3658. *Journal of Economic Perspectives* is owned, managed, and published by the American Economic Association, a nonprofit educational organization, located at 2014 Broadway, Suite 305, Nashville, Davidson County, TN 37203. The Editor is Heidi Williams. The Managing Editor is Timothy Taylor: *Journal of Economic Perspectives*, 2403 Sidney Street, Suite 260, Pittsburgh, PA 15203. The tax status of the American Economic Association has not changed during the preceding twelve months. During the preceding twelve months, the average number of copies printed for each issue was 3,657; the average total paid and/or requested circulation, 3,181; the average total non-requested distribution, 15; the average number of copies not distributed, 476; the average total distribution, 3,181. Corresponding figures for May 2021, the issue published nearest to filing date: total number of copies printed, 3,360; total paid and/or requested circulation, 3,063; total non-requested distribution, 12; number of copies not distributed, 297; total distribution, 3,063. During the preceding twelve months, the average number of requested and paid electronic copies each issue was 916; the total average requested and paid print and electronic copies, 4,097. Corresponding figures for May 2021, the issue published nearest to filing date: number of requested and paid electronic copies, 915; the total requested and paid print and electronic copies, 3,978. Certified by Barbara Fiser, Director of Finance and Administration.

The Journal of

Economic Perspectives

Fall 2021, Volume 35, Number 4

Symposia

Criminal Justice

Geographic Disparities in Health

Articles

Feature

AMERICAN ECONOMIC ASSOCIATION